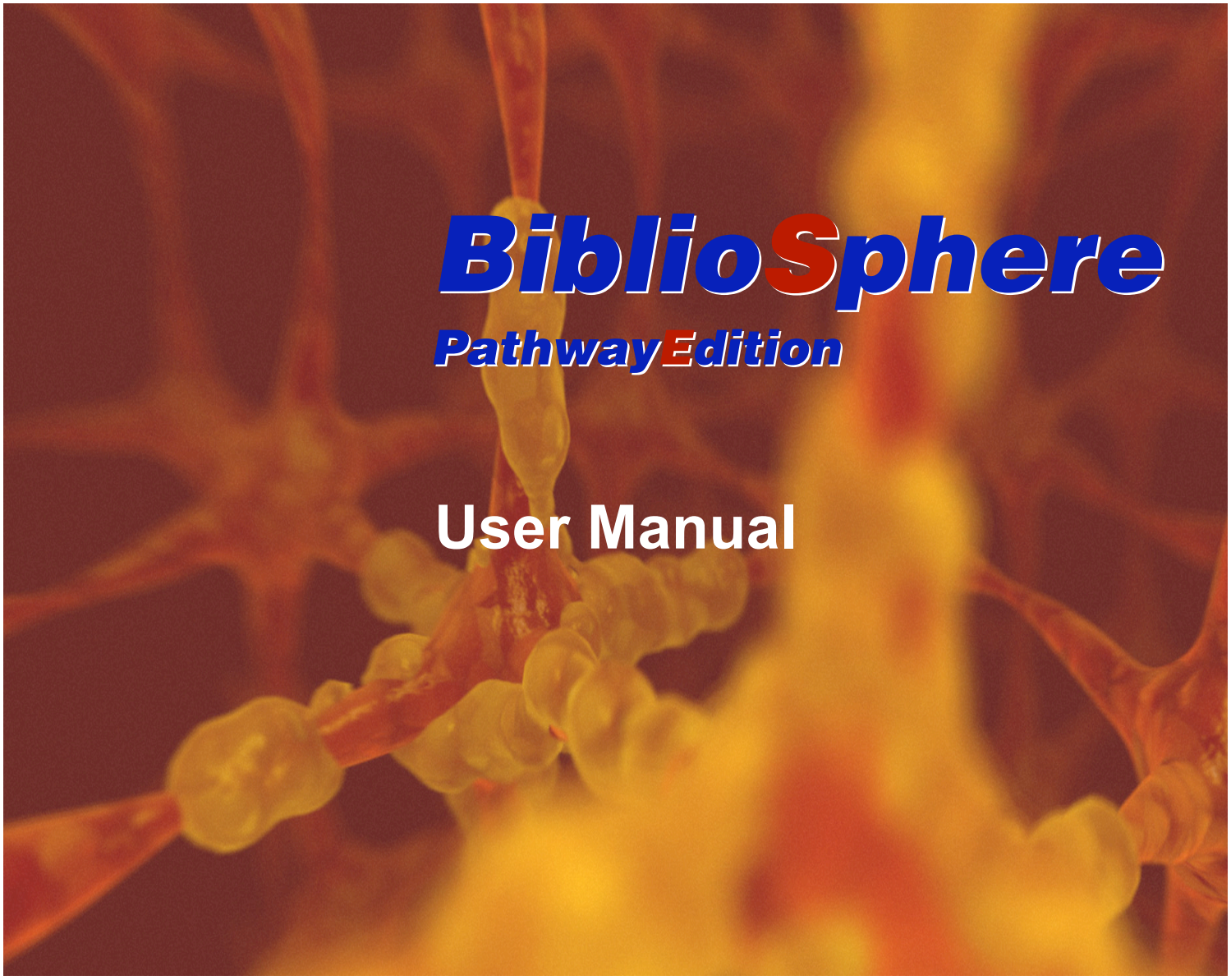
Genomatix
understanding gene regulation



© 2007 Genomatix Software GmbH

For more information please contact:

Genomatix Software GmbH
Bayerstr. 85a
80335 Munich
Germany

Phone: +49 89 599766 0
Fax: +49 89 599766 55
Email: info@genomatix.de
WWW: http://www.genomatix.de

Table of Contents

Table of Contents 2
Introduction to BiblioSphere PathwayEdition 4
Knowledge Database 5
What Data is BSPE Based on? 5
Methods and Data Sources 6
Network Generation and Analysis 8
Data Input, Network Calculation and Biological Ranking 8
User Controlled Network Construction and Analysis 9
Technical Requirements 12
Operating Systems 12
Java Runtime Environment 12
Connection to the Internet 14
Installation of BSPE 14
Download 14
Get Login and Password 16
Registration 16
Change Password 18
Password Policy 19
Installation 20
Configuration of BSPE 22
Proxy Configuration 24
SSL Configuration 25
Server Configuration 26
Check for Updates 27
Turning on Automatic Update Notification 28
Updating your Application Manually 29
How to Prepare your Input Data 31
Gene Identifiers 31
Gene List 31
User Interface 31
Excel Files 32
Starting an Analysis 32
Inputting your Data 32
Sign In 32
Project Management 33
Create a New Project 33
Edit an Existing Project 34
Create a New Analysis 34
Input Data 35
Accessing your Analyses 38
Running your Analysis 38
Ambiguities 38
Hit List 39
Network View 41
Pathway View 41
Overview 41
Connection Modes 43
Relation Info Panel 43
Node Info Panel 44
Docking and Undocking 46
Zoom 46
Shortest Path 46
Layout Optimization 46
Color Scheme Chooser 46
Export Networks 47
Metabolic & Signal Transduction Pathways 47

© 2007 Genomatix Software GmbH

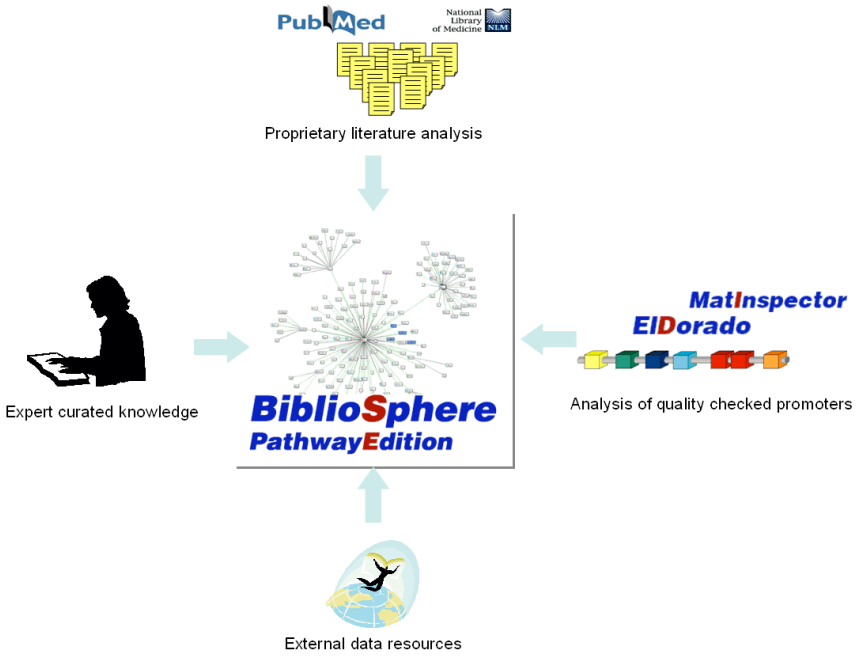
Importing your Own Annotations 48
Network Customization 52
3D View 52
Overview 52
Information about Genes and Connections 54
Focus on Gene Subnets (Clusters) 54
Co-Citation Browser 55
Link to PubMed 55
Tagged Sentences 55
Table Views 56
Overview 56
Documents Table 56
TF Analysis 57
Genes 58
Gene-Gene Connections 59
Cellular Component View 60
Protocol Panel 61
Status Bar 61
Network Filtering 62
Overview 62
Filter Panel 62
Literature Analysis Filter 62
Co-Citation Filter 62
Free Text Filter 65
Biological Entity Filter 65
Overview 65
Gene Ontology Filter 66
MeSH Filter 67
Tissue Filter 69
User Data Filter 70
Sub Network Filter 70
Statistical Analysis 71
Statistical Rating 71
Superimposition of Filters 71
BiblioSphere PathwayEdition Help 72
Online Resources 72
Contacting Genomatix 72
Glossary 72
Literature 72

© 2007 Genomatix Software GmbH

Introduction to BiblioSphere PathwayEdition

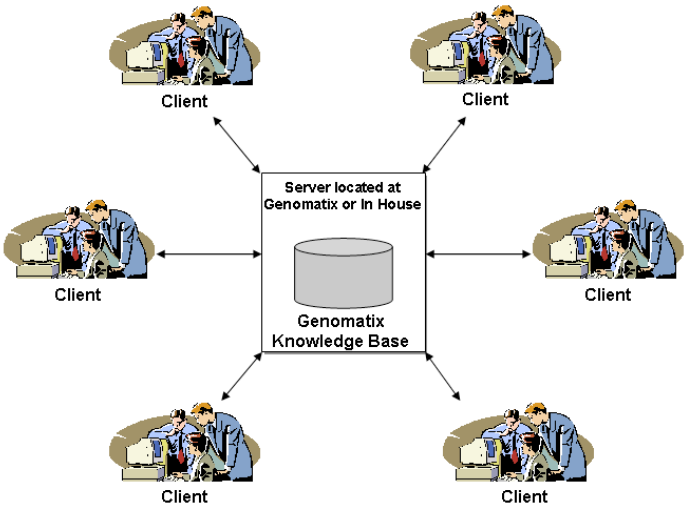
BiblioSphere PathwayEdition (BSPE) is a next-generation software system for dynamic, data driven retrieval and analysis of gene relation networks.

BSPE is the only system available which combines literature analysis with proprietary genome annotation and promoter analysis. Relations between biological entities are based on **independent information sources** (**multiple lines of evidence**) which provides insights beyond current literature knowledge.



BSPE is the only application where the user starts his analysis on base of the **entire network** of input genes, correlated genes and their biological connections. Various tools facilitate focusing on the most relevant biological context.

BSPE is based on a client-server architecture which includes the BiblioSphere PathwayEdition Knowledge Database (BSPEKD) on the server side and a retrieval, visualization and analysis system for Network Generation and Analysis which is installed as a stand alone tool on the user's computer (client side).



Knowledge Database

The BiblioSphere Pathway Knowledge Database (BSPEKD) is a one-of-a-kind structured resource of gene identifiers and relationships between biological entities. Relationships are created from more than 15 million PubMed abstracts **plus** the analysis of Genomatix's world's largest quality checked promoter database for transcription factor binding sites with MatInspector. Ontologies, taxonomies and thesauri allow for dynamic superimposition to focus on the biological context, relevant for your research.

What Data is BSPE Based on?

The primary source of BSPE data is NCBI PubMed. This collection of over 16 million scientific abstracts is analyzed for co-citations of quality checked gene names, synonyms & relation concepts.
The Genomatix collection of gene names and synonyms is composed of gene names and synonyms supplied by NCBI Entrez Gene, checked for ambiguities by automated computational analysis and enhanced, amended and filtered by manual curation.
Co-cited genes are additionally analyzed with Genomatix MatInspector for transcription factor binding sites in their promoters.
This compilation of gene-gene connections can be filtered for gene- or document-based annotations, and checked for overrepresented features by statistical analysis. Gene based annotations include Gene Ontology analysis from Entrez Gene and information on tissue specific expression from UniGene. Document based annotations utilize MeSH annotations of abstracts supplied by PubMed and a full text index of the abstracts that make up your BSPE.
The available hand-annotated information on gene-gene connections has been assembled by Genomatix experts or is based on interaction information in Molecular Connections' NetPro™ databse.
Genes that are known to belong to a certain metabolic or signal transduction pathway are labelled in the Pathway View. The pathways are the Genomatix signal transduction pathways, and metabolic pathway associations from BioCyc.

Methods and Data Sources

PubMed

PubMed is a service of the National Library of Medicine. It includes over 15 million citations for biomedical articles. These citations are from MEDLINE and additional life science journals. The database is filtered in regard to the species of interest in the analysis and then mined for gene-gene co-citations to create the BiblioSphere.

GeneOntology

The Gene Ontology controlled vocabulary is produced and maintained by the Gene Ontology (GO) Consortium. GO provides three structured networks of defined terms to describe gene product attributes. It is widely used for the annotation of genes and gene products. Gene Ontology annotations for the BiblioSphere are supplied by NCBI Entrez Gene.

MeSH

The Medical Subject Headings (MeSH) thesaurus is a controlled vocabulary used for indexing, cataloguing, and searching for biomedical and health-related information and documents. This indexing technique was introduced by the National Library of Medicine to classify and thus easily retrieve the scientific publications in medical subjects published around the world. Today MeSH contains millions of documents in fifteen main categories, of which BSPE integrates five:

- Chemicals an Drugs
- Anatomy
- Disease
- Analytical, Diagnostic and Therapeutic Techniques and Equipment
- Biological Sciences

UniGene

UniGene is an experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene is expressed. BSPE utilizes the tissue information assigned to a gene by UniGene and integrates it into its unique hierarchical Tissue Filter.

NetPro™

NetPro™ is Molecular Connections' comprehensive fully hand-curated knowledgebase of Protein-Protein, Protein-Small molecules DNA and RNA interactions, consisting of more than 200,000 interactions captured from approximately 1,400 published journals covering more than 31,000 references. BiblioSphere integrates NetPro™ in its network graphs and connection info.

STKE

STKE is an online resource devoted to the understanding of cell signalling developed by the American Association for the Advancement of Science (AAAS) and hosted by Stanford University's HighWire Press.
BiblioSphere provides links to STKE's connections maps on the web.

KEGG Pathway

KEGG (Kyoto Encyclopedia of Genes and Genomes) provides, among other bioinformatic data, manually drawn pathway maps. KEGG is maintained by the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.
BiblioSphere provides links to KEGG pathway maps on the web.

BioCarta

BioCarta provides interactive graphic models of molecular and cellular pathways as part of an open-source project.
BiblioSphere provides links to BioCarta pathway graphs on the web.

Filter statistics

BSPE's hierarchical filters use statistical analysis to check for over- or underrepresented groups of genes and abstracts.

Statistical Rating:
For each term in a hierarchical filter a statistical analysis is performed, based on the number of observed and expected annotations for terms. The z-score of this item shows whether a certain annotation, or group of annotations, is over- or underrepresented in your set of genes. This can help you to determine if an accumulation of annotations in a branch of the tree is meaningful or not.

Z-Score:
The z-score of a term indicates how far, and in what direction, that term deviates from its distribution's mean, expressed in units of its distribution's standard deviation. The general equation for the calculation of z is:

$$z = \frac{(observed - expected)}{std.deviation(observed)}$$

For filter statistics in BiblioSphere, the z-score is calculated as follows:

$$z = \frac{(r - n\frac{R}{N})}{\sqrt{n(\frac{R}{N})(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$$

where N is the total number of annotated genes, R the number of genes meeting the filter criterion, n the total number of genes in the analysed set, and r the number of genes meeting the filter criterion in the analysed set.

Promoter analysis

Promoter Analysis in BSPE is performed with Genomatix MatInspector™ on the Genomatix Promoter Database (GPD).
MatInspector is a tool that utilizes a library of matrix descriptions for **transcription factor binding sites** to locate matches in sequences of unlimited length.

A **large library** of predefined matrix descriptions for protein binding sites exists and has been tested for accuracy and suitability. Similar and/or related matrices have been grouped into **matrix families** .

MatInspector is almost as fast as a search for IUPAC strings but has been shown to produce superior results. It assigns a quality rating to matches (called matrix similarity) and thus allows quality-based filtering and selection of matches. Individually **optimized thresholds** for the matrix similarity are available for all matrices.

MatInspector has been described first by Quandt et al. (1995), and more recently by Cartharius et al. (2005).

Shortest path calculation

BSPE views use **shortest path algorithms** to calculate the optimal sub networks for all genes.

All-pairs shortest paths:
Gene networks in BiblioSphere contain very large numbers of connections between genes. Displaying all these connections in the pathway view would render it unreadable. Therefore a strategy is needed to reduce the number of displayed edges without losing relevant information. To achieve this, BiblioSphere PathwayEdition always displays only the shortest paths from the focused gene to all other genes in the pathway view. All other connections remain hidden, until the user changes the focus by double clicking a different gene in the graph.
To calculate the shortest path between gene pairs, **Dijkstra's algorithm** is used, a graph search algorithm that solves the single-source shortest path problem for a directed graph with non negative edge path costs. As some edges in BiblioSphere's Gene Networks are undirected, they are treated as two directed edges.
In BiblioSphere, the length of an edge between two genes is determined by the weighted lines of evidence (e.g. number of co-citations) supporting the connection - the more evidence the shorter the connection. However, as opposed to the road map example, it makes a difference whether a relation is direct or indirect in biological networks. As the number of "hops" between two nodes is not taken into account by the algorithm we needed to find a way to make use of this information to make sure that direct relations between two genes are always preferred over indirect connections.
To achieve this we defined minimum (min) and maximum (max) edge lengths, so that two minimum length edges are longer than one maximum length edge (min = (max /2) +1). This guarantees that regardless of the number of co-citations, a direct connection is always shorter than an indirect connection.

Network Generation and Analysis

BSPE can be applied for various research strategies, ranging from information retrieval about a single gene of interest up to the evaluation of microarray analysis results.

Data Input, Network Calculation and Biological Ranking

Depending on the application strategy, the user can access the system with the following input formats:

- Single gene
- List of genes
- List of genes plus numerical attribute for every gene, e.g. derived from
  - Expression microarray (e.g. as provided by Genomatix's ChipInspector)
  - Protein microarray
  - Gel electrophoresis
  - Protein mass spectrometry

BSPE accepts the following gene identifiers as input: gene symbols, gene names / keywords, locus link IDs and mRNA accession numbers (e.g. from RefSeq).

Moreover, PubMed IDs, MeSH terms, and free text can be supplied to indirectly retrieve the genes cited in the according journal abstracts.

As to details, please refer to chapter "How to prepare your input data".

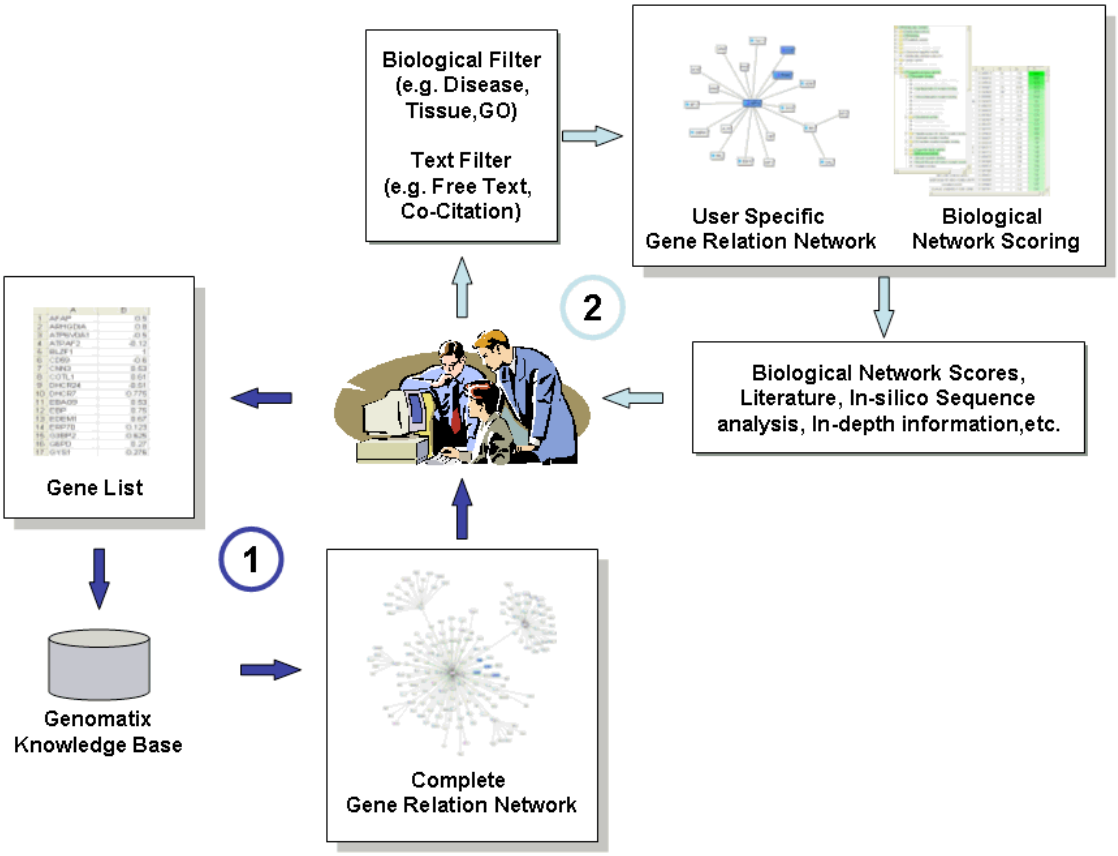
BSPE calculates the **complete** gene relation network from the list of input genes and validates and ranks pathway interactions by z-scoring on basis of the Genomatix knowledge Base. Nodes which represent input genes in the network are colour-coded if expression ratios or any other numerical attribute in the input file are provided.

User Controlled Network Construction and Analysis

Starting from the complete gene relation network, the user has full control to analyse, focus and extend the network according to a biological context of interest. For this purpose, BiblioSphere PathwayEdition offers a number of powerful tools and methods:

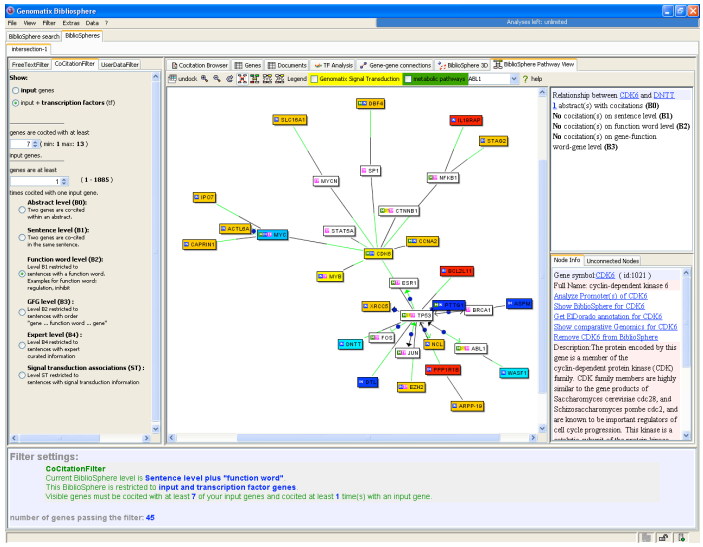
- Focus network on specific biological/experimental context
  - Unsupervised, purely data driven by "following the green path" via z-scoring
  - Supervised, by filter settings according to specific area of interest.
- Dynamic shortest path calculation and display by double clicking of central gene of interest.
- Automatic integration of pathway/network relevant transcription factors, even if those are not elements of the input set. Optional expansion by all other genes in the network beyond input list and transcription factors.
- Superimposition of additional evidence based on promoter sequence analysis
- Superimposition of additional evidence based on expert curated transcriptional regulation knowledge

The picture below shows the basic principle of the BSPE application: After an initialization step (1) which collects all information for the given set of input genes and their association to other genes, the analysis step (2) takes place – fully under control of the user and under control of a permanent biological scoring and ranking of the gene relation network.

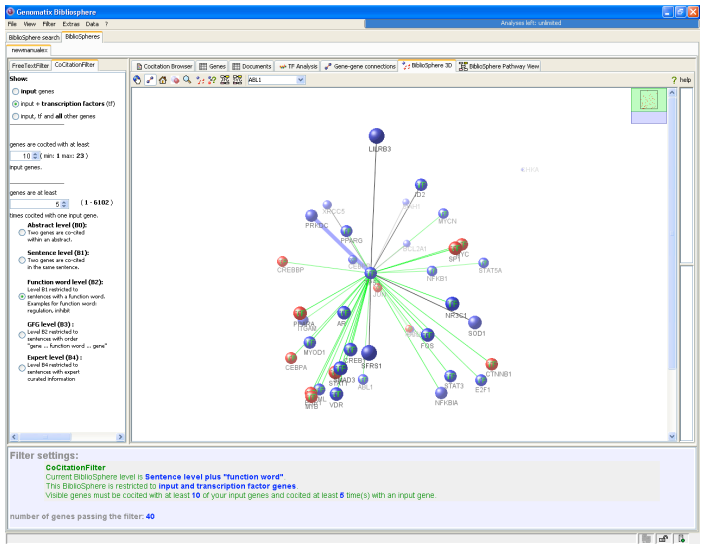


BSPE provides different kinds of views to the large amount of information that is included in the gene regulation networks:

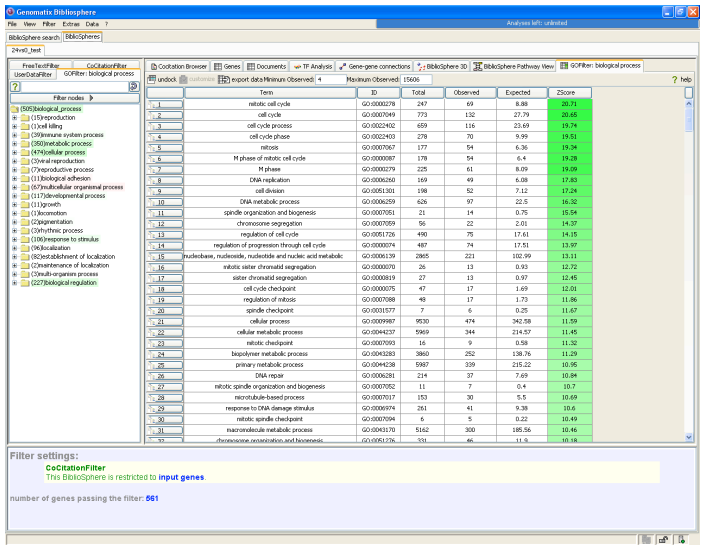
2D-view:
Application of ***shortest path algorithm*** allows focusing on different genes and their shortest path to the genes within the network. The Genomatix Knowledge base is ***continually updated*** to provide the most current data from literature and sequence analysis.



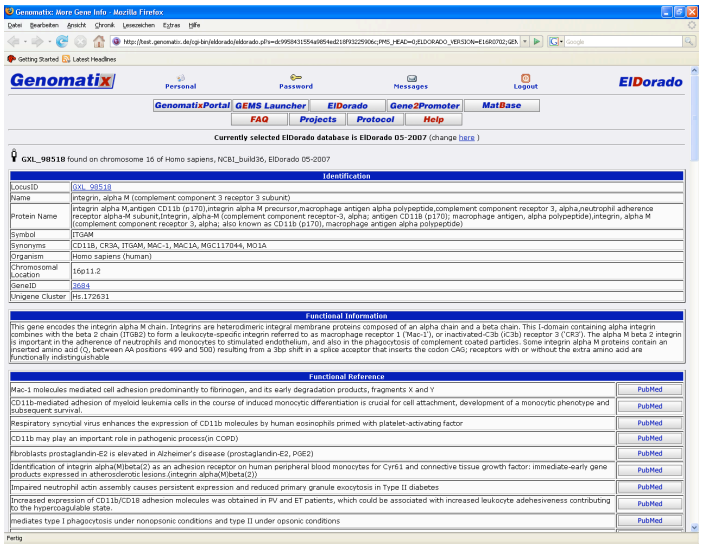
3D view:
Quick identification of ***closely related gene groups***. The distance between the entities in the 3D view reflects the number of abstracts in which the two genes are co-cited.



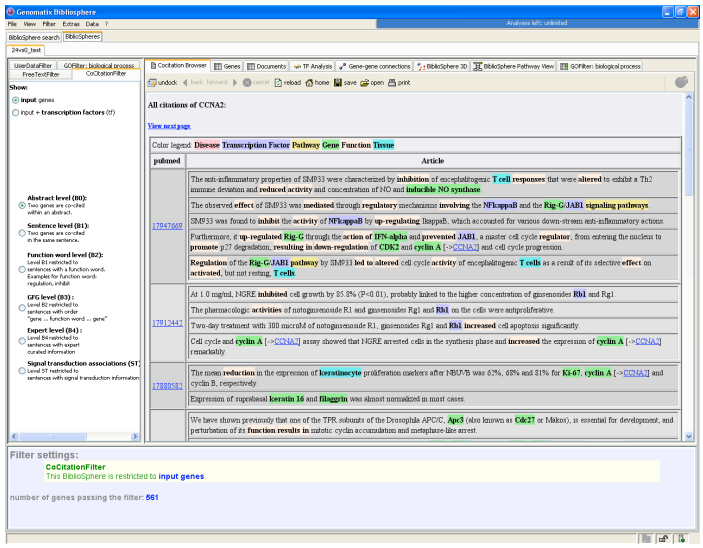
Biological Scoring:
Networks are scored according to overrepresentation of genes in ***biological categories,*** such as biological process, disease, tissue, etc. A total of 10 different categories is available.



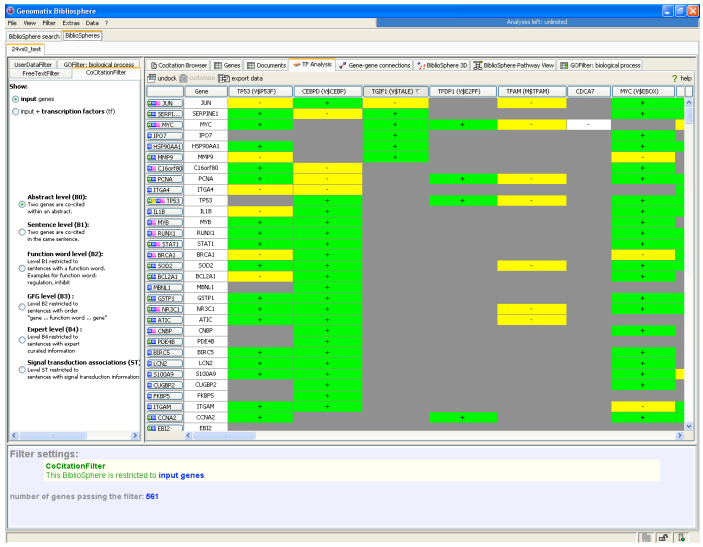
Gene Info:
An ***exhaustive summary*** provides a quick overview about a gene of interest. Gene synonyms, functional description, transcript variants, etc. are included.



Literature analysis:
Tagged sentences allow a quick overview of the relevant sentences of an abstract.



Promoter analysis:
Quick identification of transcription factor – gene relations defined by ***binding sites on promoter level***.



Technical Requirements

The following chapter explains the technical requirements to install the BiblioSphere PathwayEdition client application on your computer.

Operating Systems

The application is certified to run under the following operation systems:

Windows systems:

- Windows 98, SE, 2000, ME, XP
- 80 MB hard disc space
- 512 MB RAM recommended

Macintosh systems:

- At least MacOS 10.3
- 80 MB hard disc space
- 512 MB RAM recommended

Linux/Unix systems:

- 80 MB hard disc space
- 512 MB RAM recommended

If you do not have any of these operation systems or if you are not sure about your operation system, please contact the Genomatix customer support (support@genomatix.de)

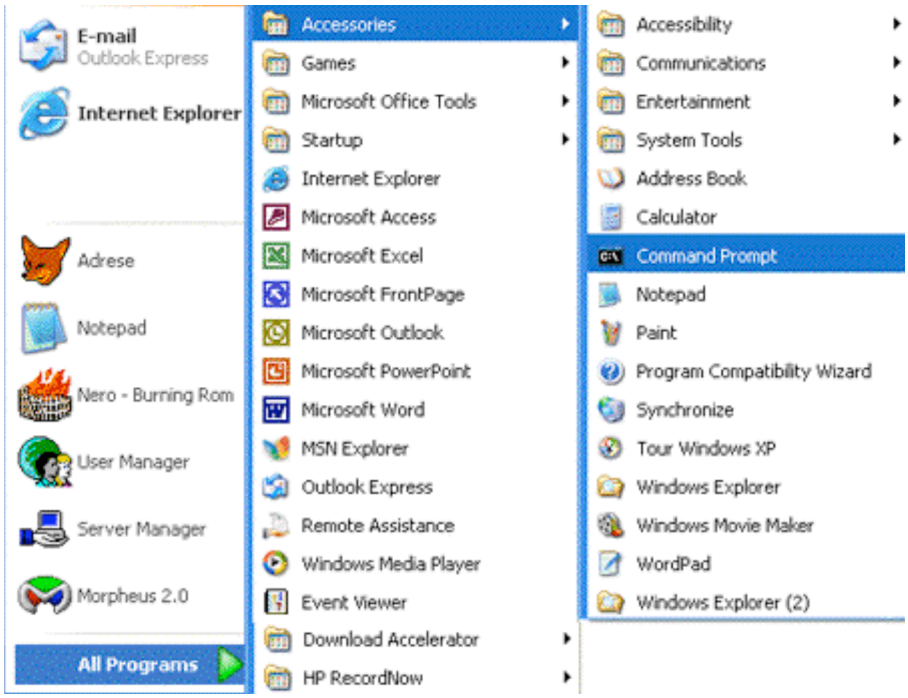
Java Runtime Environment

In order to run the BSPE application, you will need **Java 1.5** or higher.

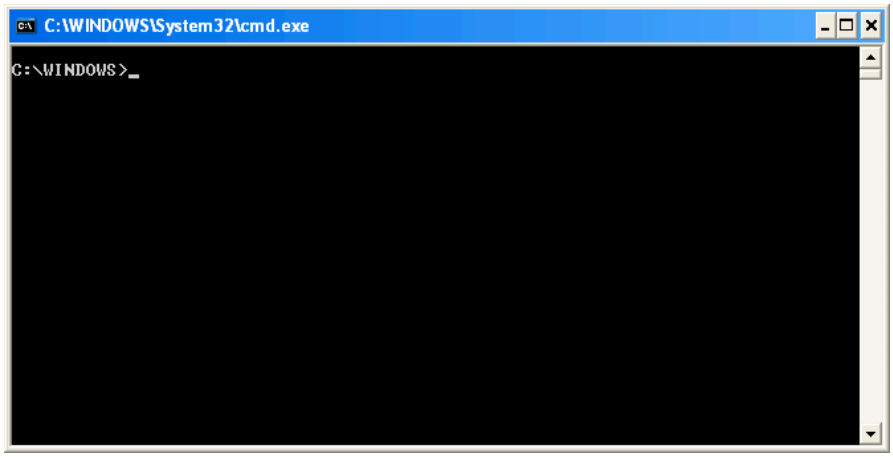
To test if you have an appropriate Java version already installed on your system, type "**java –version**" on command line.

Here is an example for windows users how to check the installed java version:

Click on Start/All Programs/Accessories/Command Prompt (see screenshot below).



A command window will pop up:



Type in ***java –version*** and press Enter

If Java is installed, you will get an output like

```
C:\>java -version
java version "1.5.0_06"
Java(TM) 2 Runtime Environment, Standard Edition (build 1.5.0_06-b05)
Java HotSpot(TM) Client VM (build 1.5.0_06-b05, mixed mode, sharing)
```

If Java is **not** yet installed on your computer, or if you have a Java version older than 1.5, please follow the link http://www.java.com/ to download and install the newest version of Java (at least version 1.5).

Connection to the Internet

BSPE retrieves information from the BSPEKD Database which is hosted by Genomatix. Therefore an internet connection is needed to run BSPE.

Alternatively, BSPEKD can be installed on a server at your site (please contact support@genomatix.de for details on in house installations). In this case an intranet connection to your server would be required.

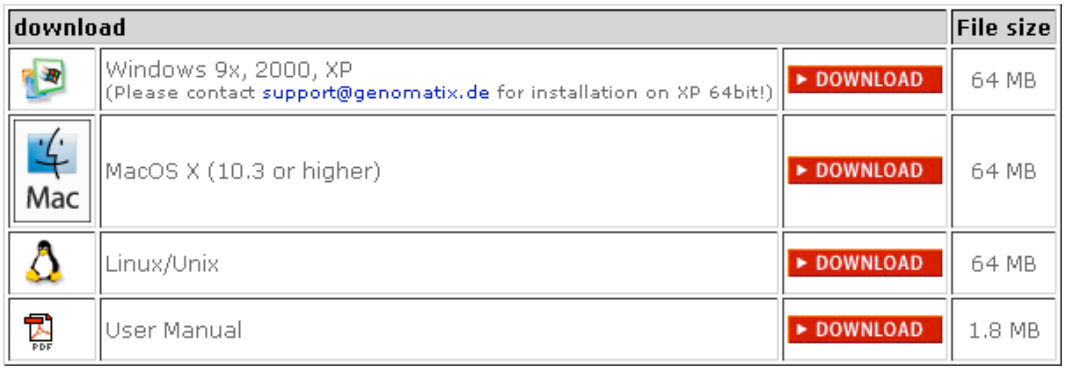
Installation of BSPE

BSPE is a Java program which must be installed locally on your computer. Please proceed for download and installation as follows.

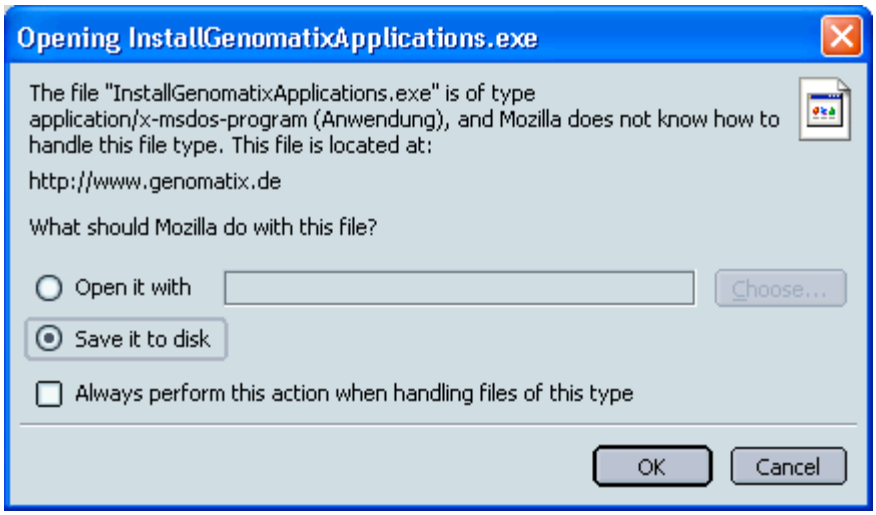
Download

To download BSPE, please follow the following steps:

1. Create a folder on you hard disk where you want to store the installer
2. Switch to http://www.genomatix.de/products/BiblioSphere/BiblioSpherePE5.html
3. Choose your operating system from the download
4. Click on the download button next to your operating system



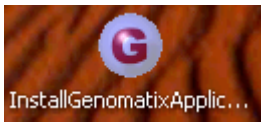
Clicking on the download-icon will result in the following screen:


Opening InstallGenomatixApplications.exe
The file "InstallGenomatixApplications.exe" is of type application/x-msdos-program (Anwendung), and Mozilla does not know how to handle this file type. This file is located at:
http://www.genomatix.de
What should Mozilla do with this file?
Open it with
Choose...
Save it to disk
Always perform this action when handling files of this type
OK
Cancel

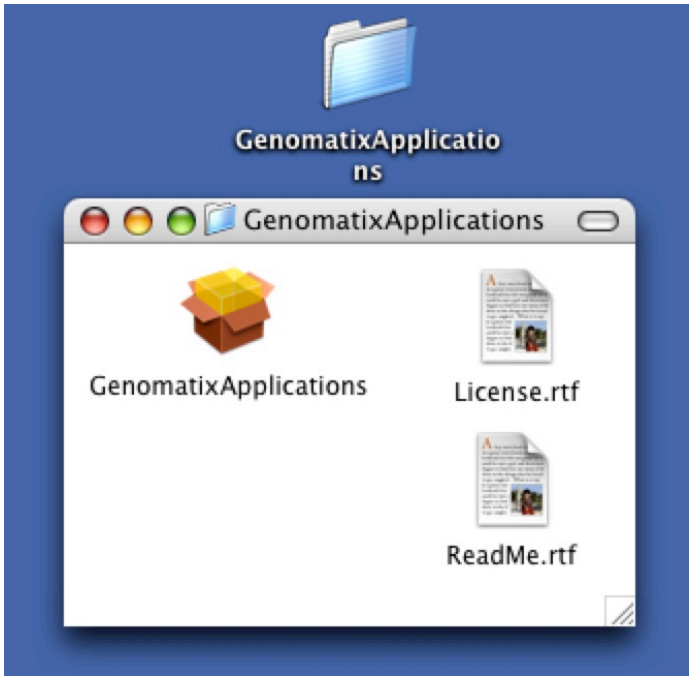
Choose the option “save to disk” and click “ok”

A window will show up, where you can choose a folder to save the file. Choose the folder where you would like to save the installer and press ok.

If the installer is successfully downloaded, Windows users should see the following icon with the subtitle “InstallGenomatixApplication.exe”



Mac users will find a folder named "GenomatixApplications" on their desktop or in their designated download folder. It contains an installer package, a ReadMe and the license file. Double clicking the "GenomatixApplications" installer package will start the installation of the software.



Get Login and Password

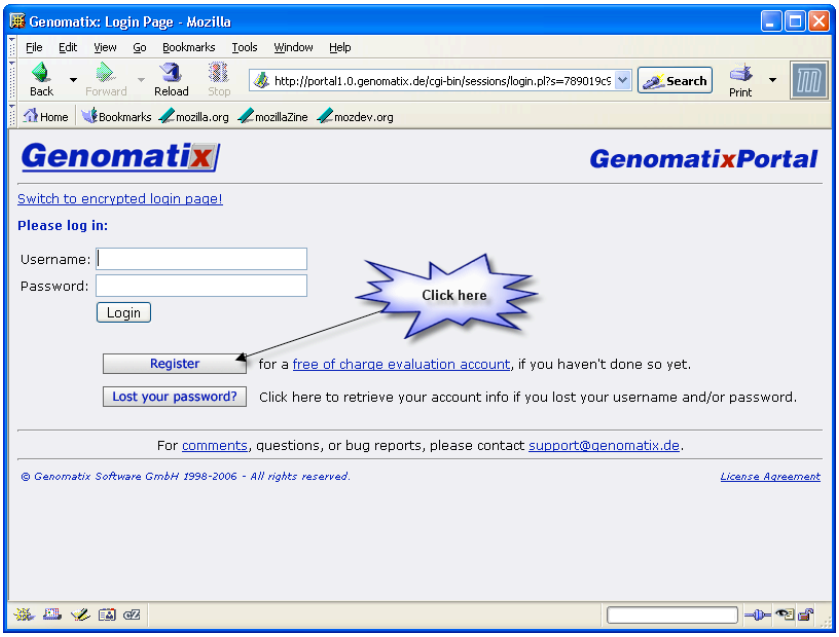
To apply the BSPE application you need a login and a password. Registration is **free of charge**. An e-mail with your personal username and password will be sent to you right away.

Registration

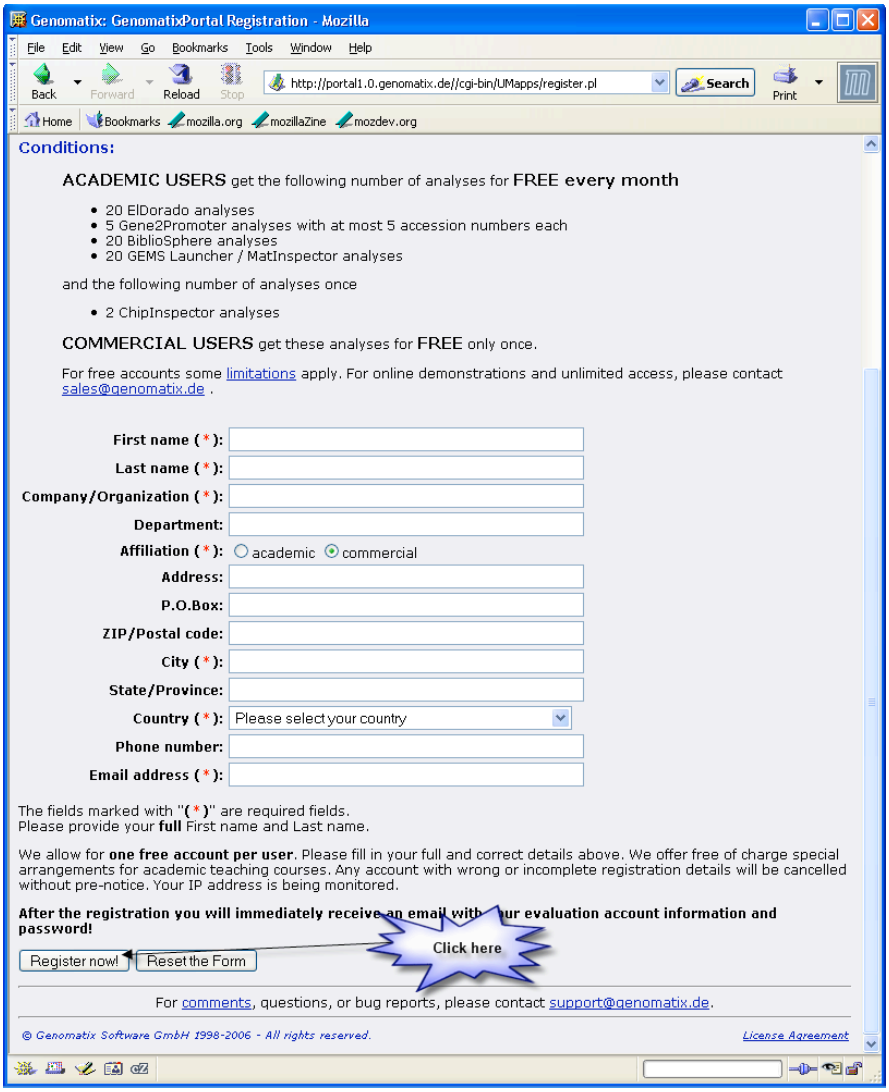
Open your internet browser and switch to www.genomatix.de.
Click on “Login” in the navigation panel of the webpage.

Genomatix
understanding gene regulation
Home • Sitemap • Contact
Software
Collaborative Research
Training
Customers
About us
Downloads
News & Events
Careers
Login
Get a free evaluation account
Buy a license online
Search our site:
Click here
Software
Use our tools to understand the molecular mechanisms of gene regulation as a central part of systems biology. Our programs will provide comprehensive, reliable results based on scientific algorithms.
Collaborative & Contract Research
More than ten years of scientific experience make Genomatix an excellent partner for all projects involving gene regulation, regulatory pathways or microarray analysis. Learn more about our recent collaborations from our project list or contact us at research@genomatix.de.
Training
To help you get the most from our products, Genomatix offers various training options including step-by-step tutorials and online webmeetings accessible directly from your workplace as well as tutored seminars at our headquarters in Munich or even at your site.
Genomatix releases new versions of ChipInspector and BiblioSphere
Genomatix has released ChipInspector 1.3.5, with updated probe annotation and a new position based analysis feature, which allows to export significant regions. BiblioSphere PathwayEdition 7.0 now contains data from MolecularConnection's NetPro™ database providing an additional line of expert curated evidence and a new filter method.
Both applictions are available from our download area.
Genomatix Expands Operations in North America
A new subsidiary and training academy have been incorporated in Ann Arbor, Michigan, USA.
Read more in our press release.
Training Resources
read manuals or tutorials
book a training seminar
Currently we offer courses on promoter and microarray analysis. To learn more about their contents have a look at the overview page or contact us at seminars@genomatix.de.
© 1998-2007 Genomatix Software GmbH - All rights reserved
Corporate Information • Privacy Policy • Trademarks

If you do not have an account yet, please click on “Register”.


Genomatix: Login Page - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://portal1.0.genomatix.de/cgi-bin/sessions/login.pl?s=789019c9
Search Print
Home Bookmarks mozilla.org mozillaZine mozdev.org
Genomatix
GenomatixPortal
Switch to encrypted login page!
Please log in:
Username:
Password:
Login
Click here
Register
for a free of charge evaluation account, if you haven't done so yet.
Lost your password?
Click here to retrieve your account info if you lost your username and/or password.
For comments, questions, or bug reports, please contact support@genomatix.de.
© Genomatix Software GmbH 1998-2006 - All rights reserved.
License Agreement

Fill in the form – please enter your e-mail correctly.


Genomatix: GenomatixPortal Registration - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://portal1.0.genomatix.de//cgi-bin/UMapps/register.pl
Search
Print
Home Bookmarks mozilla.org mozillaZine mozdev.org
Conditions:
ACADEMIC USERS get the following number of analyses for FREE every month
20 ElDorado analyses
5 Gene2Promoter analyses with at most 5 accession numbers each
20 BiblioSphere analyses
20 GEMS Launcher / MatInspector analyses
and the following number of analyses once
2 ChipInspector analyses
COMMERCIAL USERS get these analyses for FREE only once.
For free accounts some limitations apply. For online demonstrations and unlimited access, please contact sales@genomatix.de .
First name (*):
Last name (*):
Company/Organization (*):
Department:
Affiliation (*): academic commercial
Address:
P.O.Box:
ZIP/Postal code:
City (*):
State/Province:
Country (*): Please select your country
Phone number:
Email address (*):
The fields marked with "(*)" are required fields.
Please provide your full First name and Last name.
We allow for one free account per user. Please fill in your full and correct details above. We offer free of charge special arrangements for academic teaching courses. Any account with wrong or incomplete registration details will be cancelled without pre-notice. Your IP address is being monitored.
After the registration you will immediately receive an email with your evaluation account information and password!
Register now!
Reset the Form
Click here
For comments, questions, or bug reports, please contact support@genomatix.de.
© Genomatix Software GmbH 1998-2006 - All rights reserved.
License Agreement

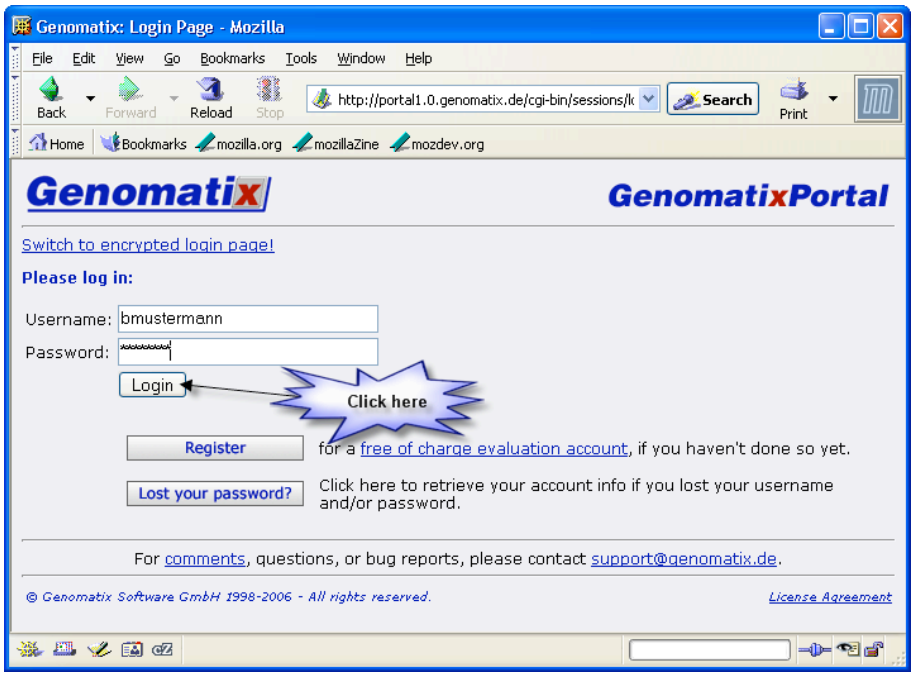
Check your e-mail. A mail with you login data should be sent to you right away.

```
Dear Bernd Mustermann,


your registration for a Genomatix evaluation account was successful!
Here are your account details:

      login name: bmustermann
        password: nUDbjPnW

Please note that username AND password are case-sensitive.

You will find the GenomatixSuite at
   http://portal1.0.genomatix.de/cgi-bin/./eldorado/main.pl
You can change your password using the "change password" link
at the top of the program page or on your "Personal profile" page.

You can also subscribe to unlimited MatInspector access on your
"Personal profile" page.

If you should run into trouble logging onto our server, please first
read
http://portal1.0.genomatix.de/online_help/help/techfaq.html

After that do not hesitate to contact us in case of any problems or
questions.


Best regards,

Your Genomatix support team


~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
For comments or suggestions please contact
          Genomatix Software
      at support@genomatix.de !
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

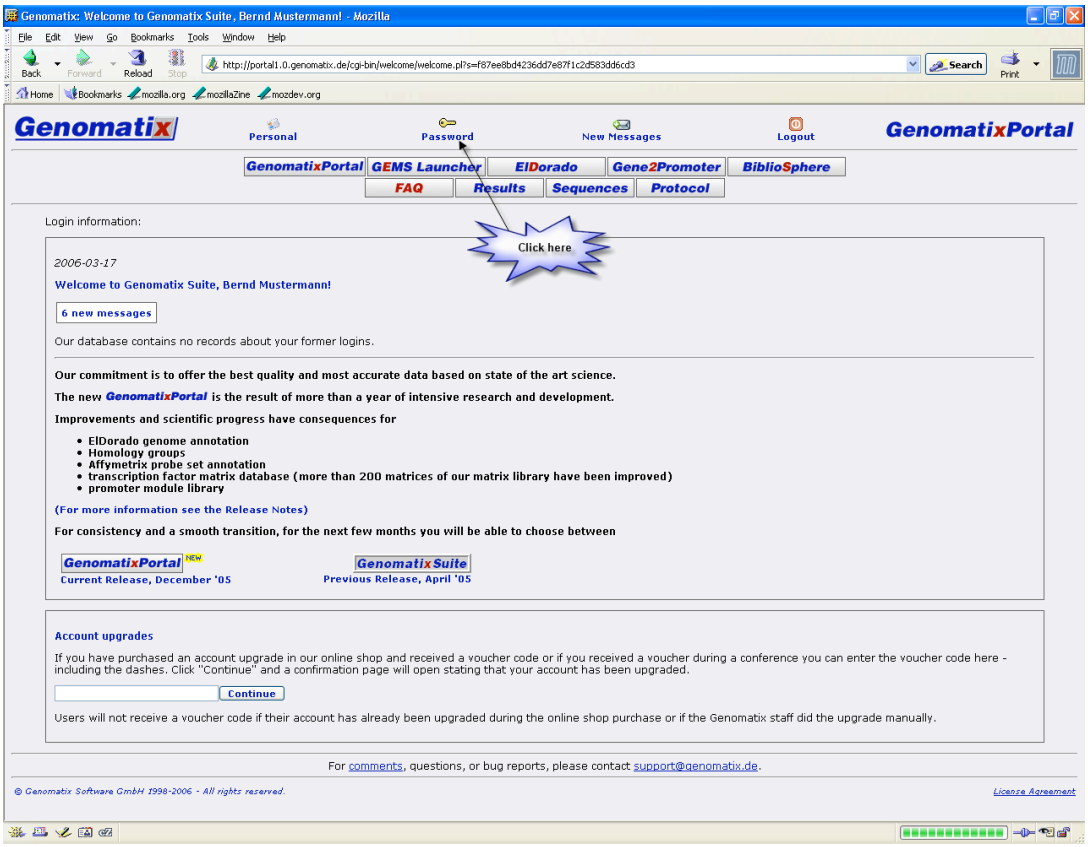
The login and password is not only valid for BiblioSphere PathwayEdition but for all Genomatix products.

Change Password

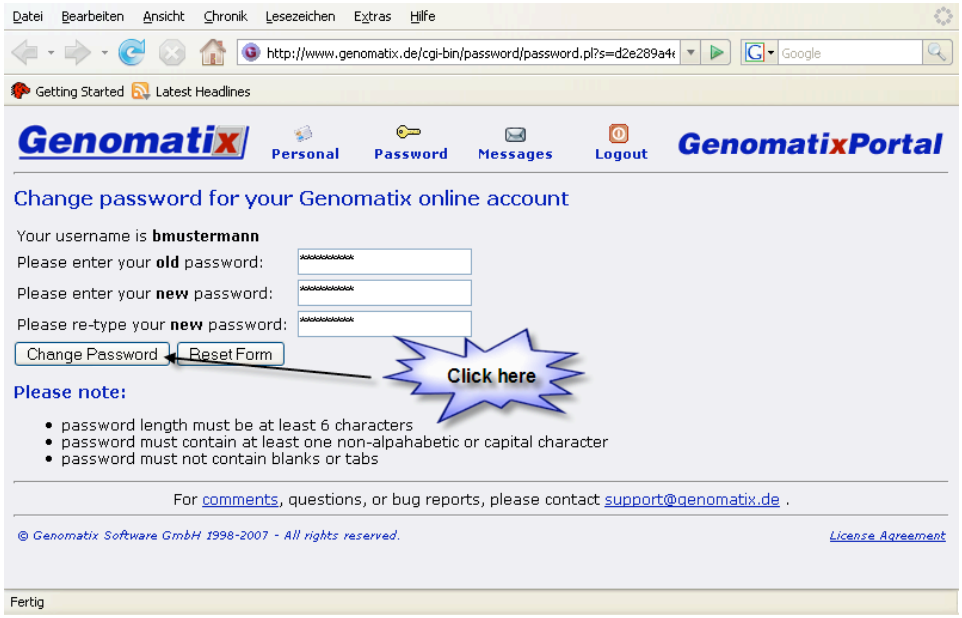
Open your internet Browser and switch to www.genomatix.de.
Click on “Login” in the upper right corner of the webpage (see above)
Enter your login and password which was sent to you via e-mail.


Genomatix: Login Page - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://portal1.0.genomatix.de/cgi-bin/sessions/lo
Search
Print
Home Bookmarks mozilla.org mozillaZine mozdev.org
Genomatix
GenomatixPortal
Switch to encrypted login page!
Please log in:
Username: bmustermann
Password:
Login
Click here
Register
for a free of charge evaluation account, if you haven't done so yet.
Lost your password?
Click here to retrieve your account info if you lost your username and/or password.
For comments, questions, or bug reports, please contact support@genomatix.de.
© Genomatix Software GmbH 1998-2006 - All rights reserved.
License Agreement

After login you will see the following page. Click on “Password”.


Genomatix: Welcome to Genomatix Suite, Bernd Mustermann! - Mozilla
Personal
Password
New Messages
Logout
GenomatixPortal
GenomatixPortal
GEMS Launcher
ElDorado
Gene2Promoter
BiblioSphere
FAQ
Results
Sequences
Protocol
Click here
Login information:
2006-03-17
Welcome to Genomatix Suite, Bernd Mustermann!
6 new messages
Our database contains no records about your former logins.
Our commitment is to offer the best quality and most accurate data based on state of the art science.
The new GenomatixPortal is the result of more than a year of intensive research and development.
Improvements and scientific progress have consequences for
ElDorado genome annotation
Homology groups
Affymetrix probe set annotation
transcription factor matrix database (more than 200 matrices of our matrix library have been improved)
promoter module library
(For more information see the Release Notes)
For consistency and a smooth transition, for the next few months you will be able to choose between
GenomatixPortal
Current Release, December '05
GenomatixSuite
Previous Release, April '05
Account upgrades
If you have purchased an account upgrade in our online shop and received a voucher code or if you received a voucher during a conference you can enter the voucher code here - including the dashes. Click "Continue" and a confirmation page will open stating that your account has been upgraded.
Continue
Users will not receive a voucher code if their account has already been upgraded during the online shop purchase or if the Genomatix staff did the upgrade manually.
For comments, questions, or bug reports, please contact support@genomatix.de.
© Genomatix Software GmbH 1998-2006 - All rights reserved.
License Agreement

Fill in the form and click on “Change Password” to change your password.


Personal
Password
Messages
Logout
GenomatixPortal
Change password for your Genomatix online account
Your username is bmustermann
Please enter your old password:
Please enter your new password:
Please re-type your new password:
Change Password
Reset Form
Click here
Please note:
password length must be at least 6 characters
password must contain at least one non-alpahabetic or capital character
password must not contain blanks or tabs
For comments, questions, or bug reports, please contact support@genomatix.de .
© Genomatix Software GmbH 1998-2007 - All rights reserved.
License Agreement
Fertig

Password Policy

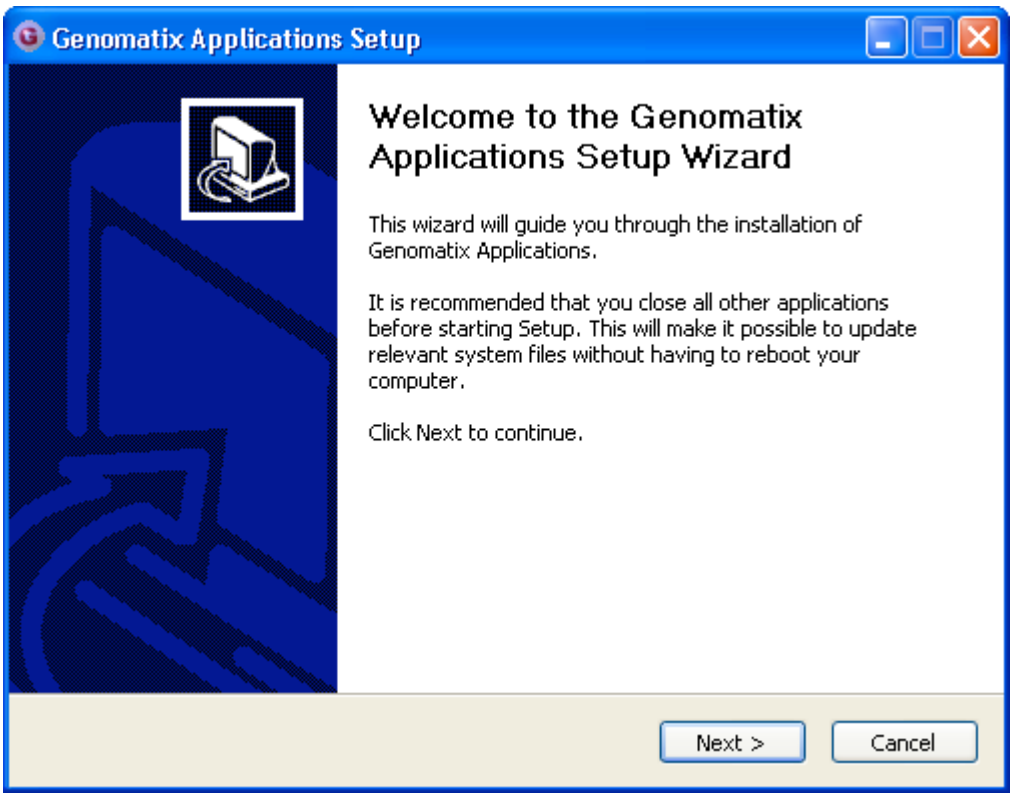
Genomatix’s password policy requires all passwords to be at least 6 characters long and must contain at least one non-alphabetic or capital character. No blanks or tabs are allowed.

Installation

Switch to the folder on your hard disk where the installer was saved. Execute the installer (see below) and follow the instructions. The installer will install both BSPE and ChipInspector.

install		
	Windows 9x, 2000, XP	doubleclick on the Genomatix install icon in the folder (requires administrator privileges!)
Mac	MacOS X (10.3 or higher)	run the GenomatixApplications installer from the downloaded disk image/folder (requires administrator privileges!)
	Linux/Unix	type **"java -jar InstallGenomatixApplications.jar"**

If you run a windows system, the following screen will pop up:

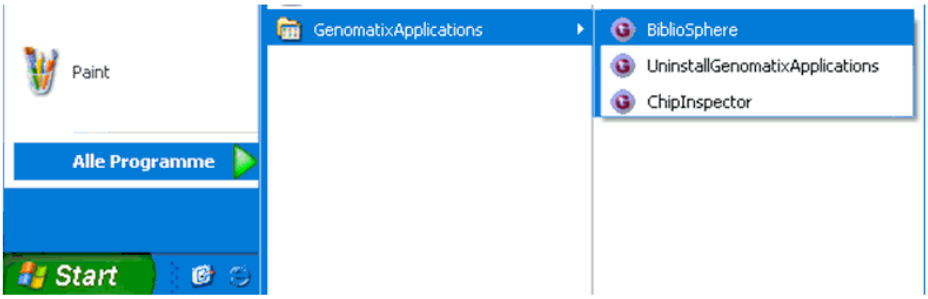

Genomatix Applications Setup
Welcome to the Genomatix Applications Setup Wizard
This wizard will guide you through the installation of Genomatix Applications.
It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.
Click Next to continue.
Next >
Cancel

Click “Next >” and follow the instructions.

After BSPE is installed successfully, you can start the application in different ways:

1. Start BSPE from the program group

After successful installation, windows users should have a new Program Group “Genomatix Applications” with an executable “BiblioSphere”. Click “Start”, ”All Programs”, ”GenomatixApplications”, ”BiblioSphere”.



2. Start BSPE from desktop

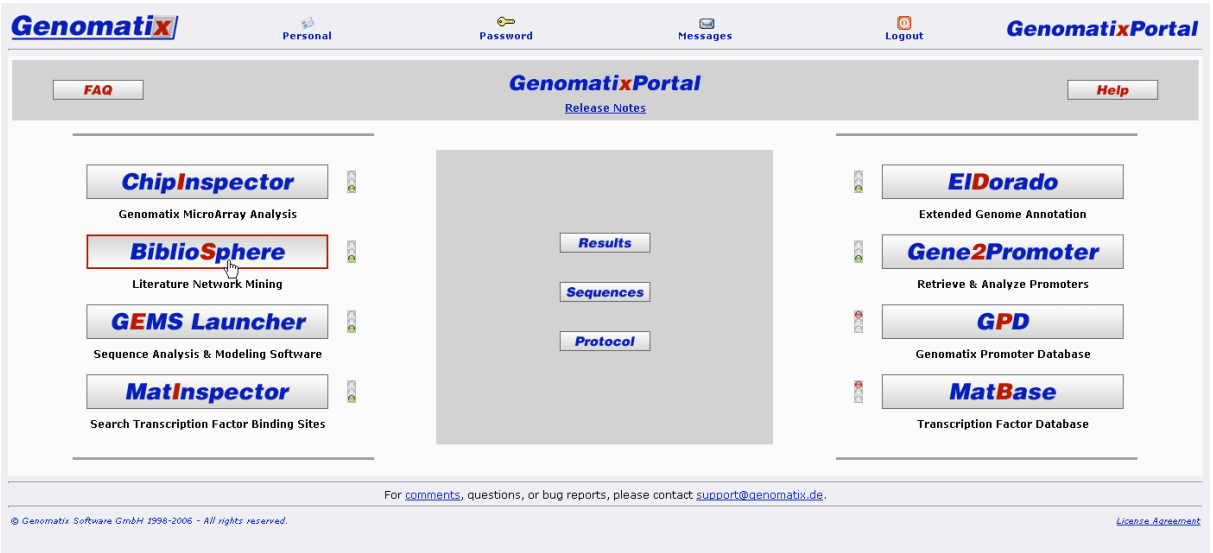
After installation you should find an Icon on your desktop:

A double click on the icon will launch the BSPE application

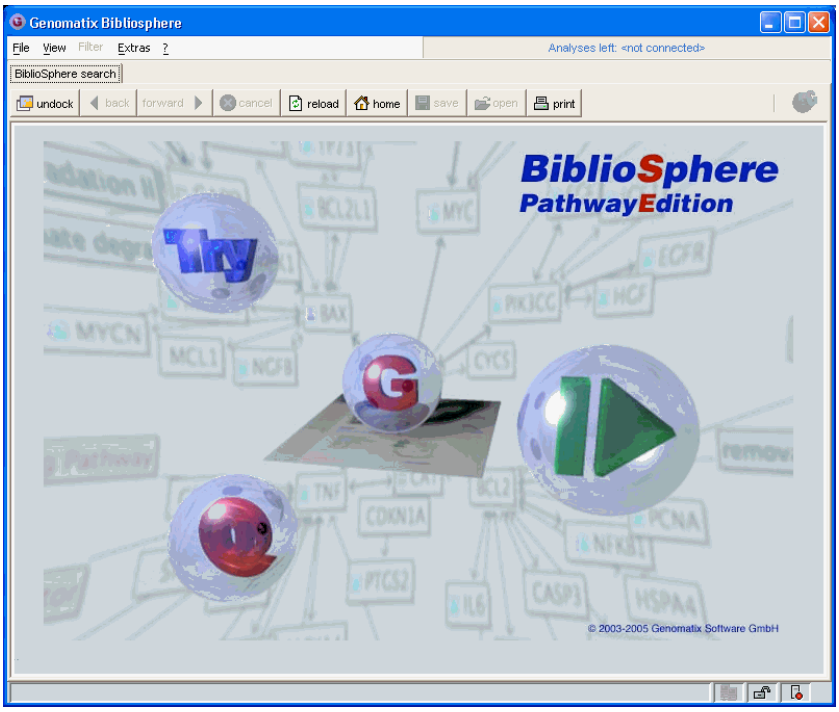
3. Start BSPE per batch file (MS Windows only)

On Windows systems, if BSPE does not start when you double click the desktop icon, you can use a batch file that you find the in a subdirectory of your Genomatix installation directory. The default location is C:\Program Files\GenomatixApplications\apps\bibliosphere\conf\bibliosphere.bat. Double click on the file in your windows explorer or, in the Windows start menu, choose “Execute…”, type in the complete file name including the path and click OK.

4. Start BSPE from the Genomatix Portal (see below)



After successful launch of the BSPE application you will see the following screen:



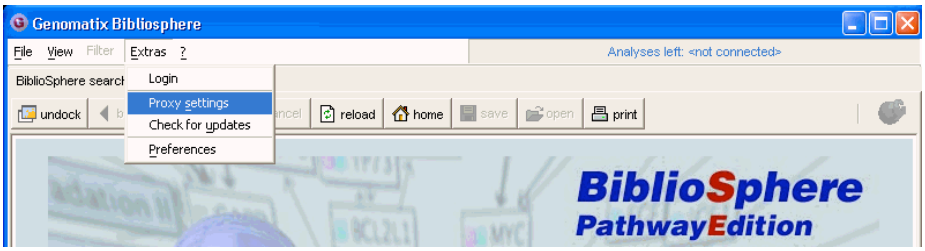
Configuration of BSPE

Before you start working with BSPE you should configure the BSPE concerning

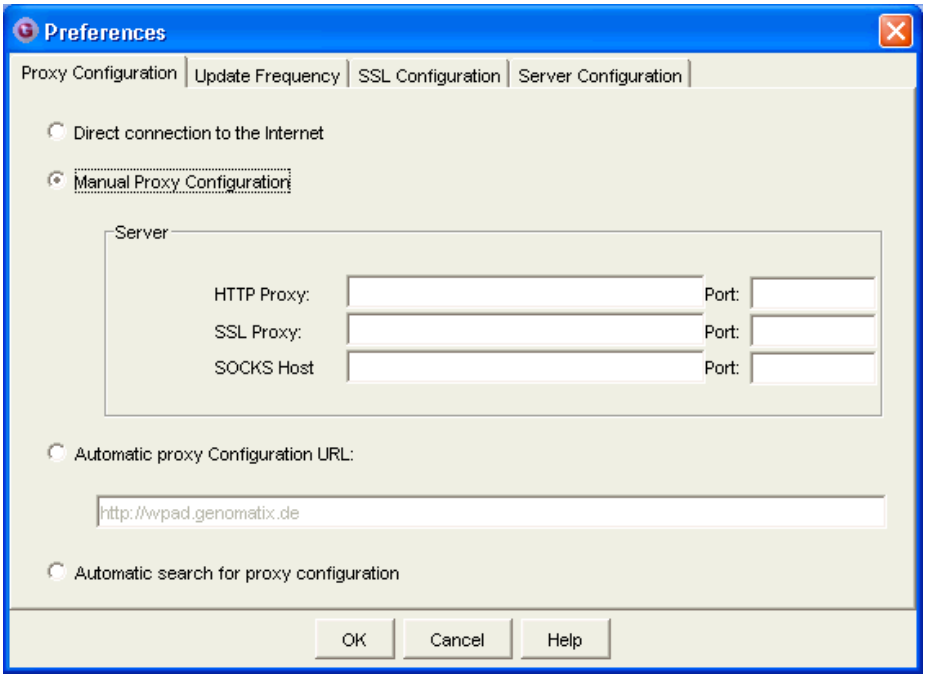
- Proxy configuration (for internet access)
- Security configuration (for secure information transfer over the internet)
- Application update (to get the latest version of BSPE online)

BSPE offers a form for configuration which can be accessed as follows:

Start BSPE application. Go to menu **"Extras"** and select "**Proxy settings**" to launch a preferences configuration dialog



You will get the following dialog which consists of four forms for the different configurations:



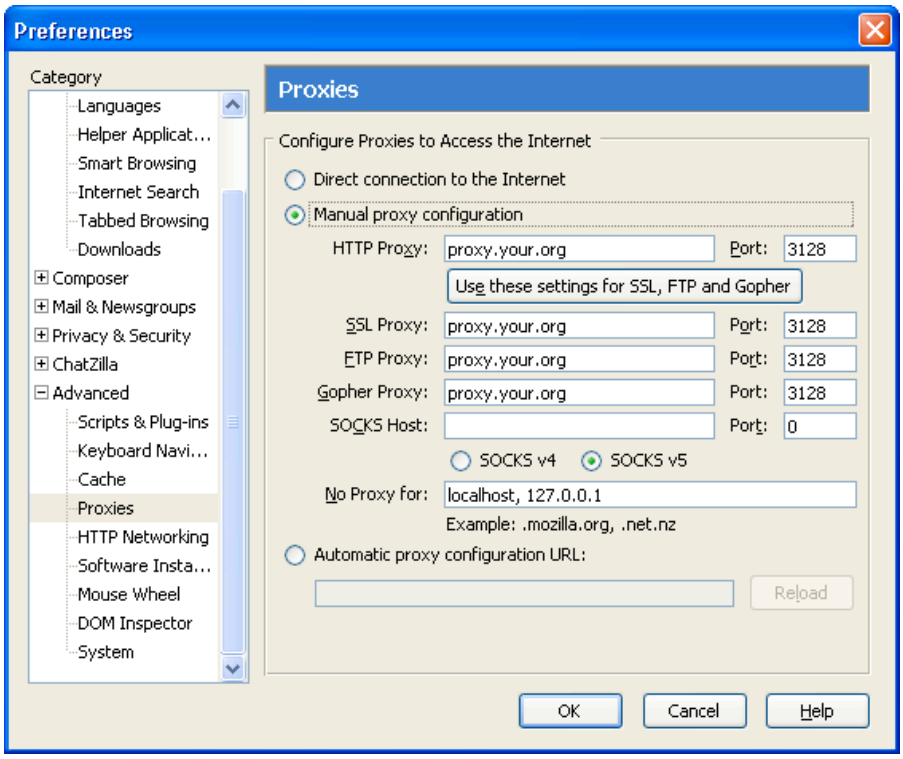
Proxy Configuration

Many companies and institutions use proxies and firewalls for secure and fast access to the Web. Thus you need to configure the BSPE application to get through your proxy or firewall.

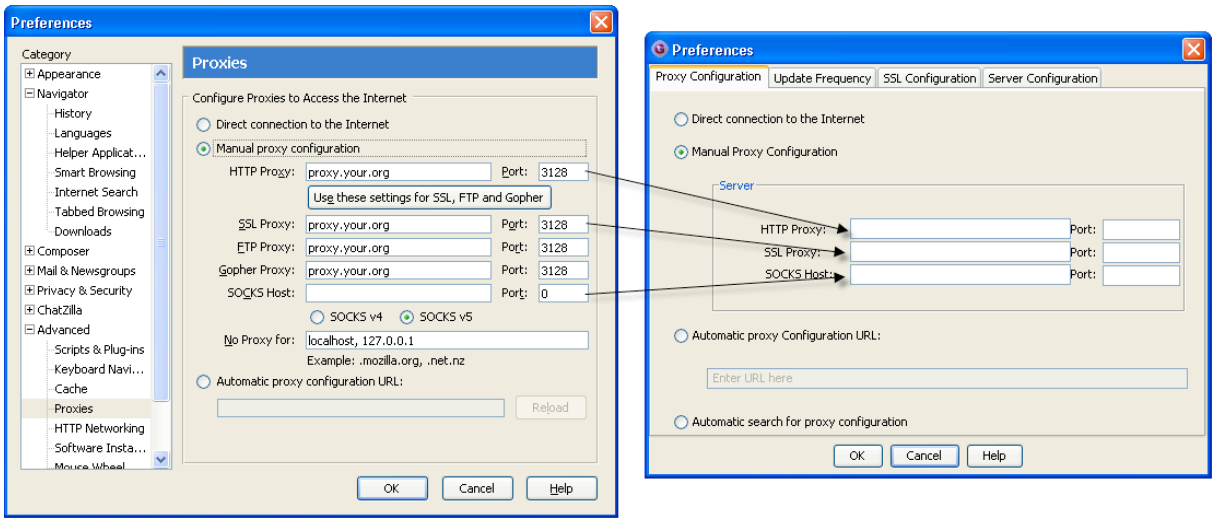
Please proceed as follows:

Get the **proxy settings** from your internet browser.
If you use internet explorer: Go to: Tools->Internet Options->Connections->**LAN settings**
If you use Netscape or Mozilla: Go to: Edit->Preferences->Advanced->**Proxies**

Below you see an example for the Mozilla browser



Configure the settings according to the configuration of your browser and press **"ok"**. Below you see an example for manual proxy configuration.



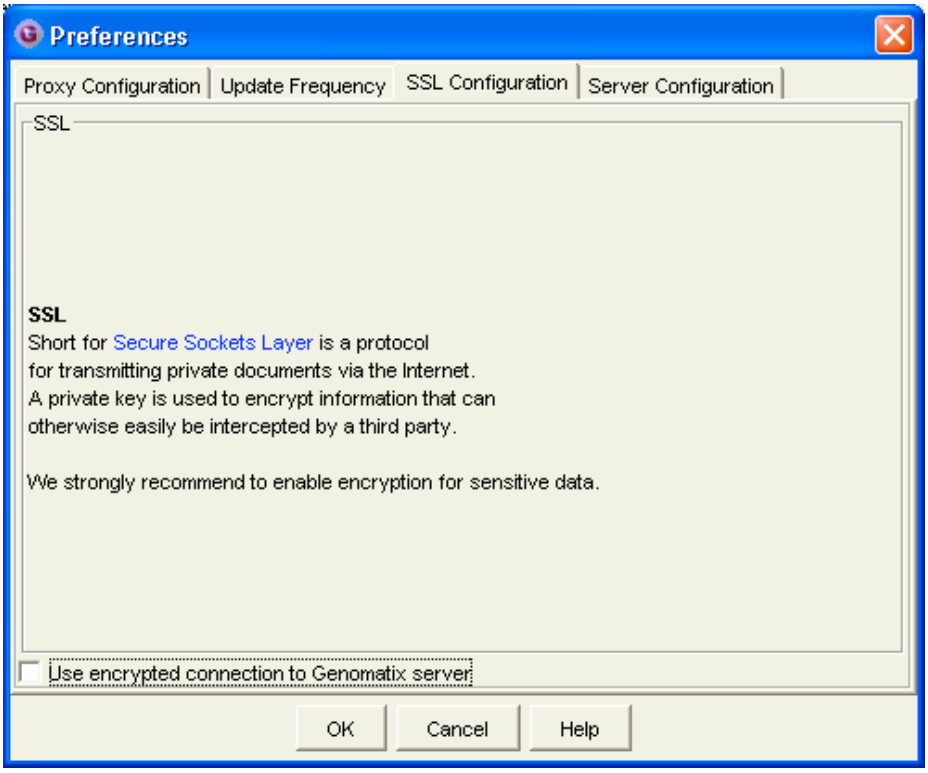
SSL Configuration

BSPE allows for encrypted communication with the server via internet via Secure Socket Layer (SSL). If you would like to use the encrypted protocol proceed as follows:

Start BSPE (see above)

Go to menu **"Extras"** and select "**Proxy settings**" to launch a preferences dialog for proxy configuration

Click on “SSL Configuration”:



Check the box next to “Use encrypted connection to Genomatix server” and then click “ok”.

If you have chosen a secure connection to the internet, a little icon will show up at the bottom of the BSPE:

Server Configuration

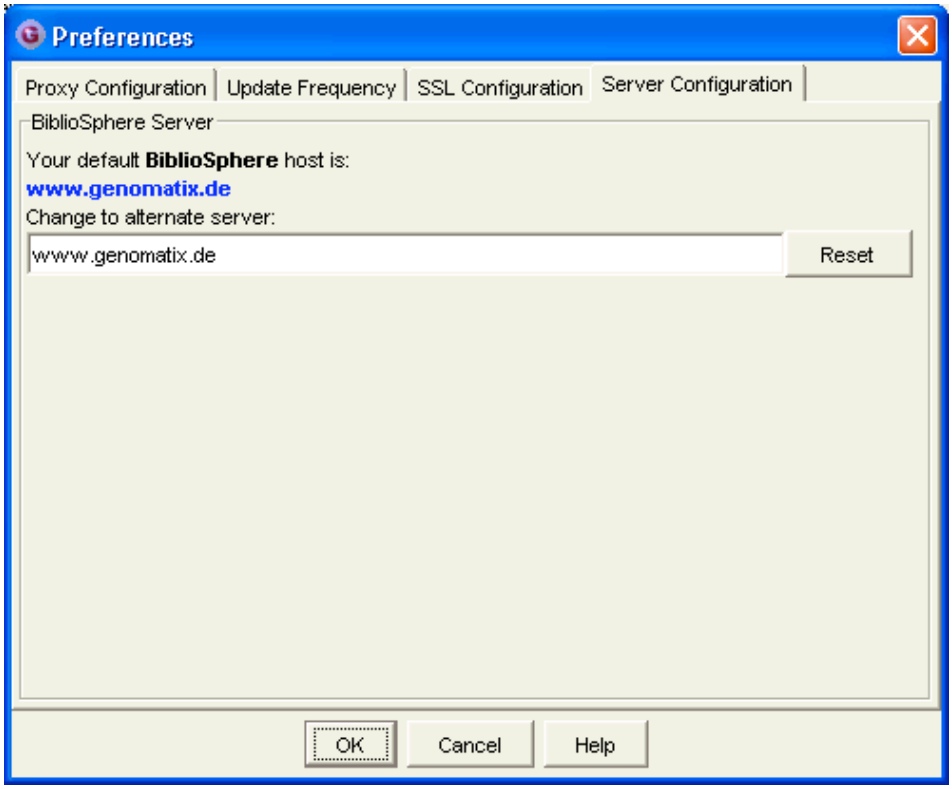
If the BSPEKD is installed in house, you will have to enter the correct server name. Please contact your system administrator. As default, the BSPEKD installed on the Genomatix server is used.

You can change the BSPEKD server as follows:

Start BSPE (see above)

Go to menu **"Extras"** and select "**Proxy settings**" to launch a preferences dialog for proxy configuration

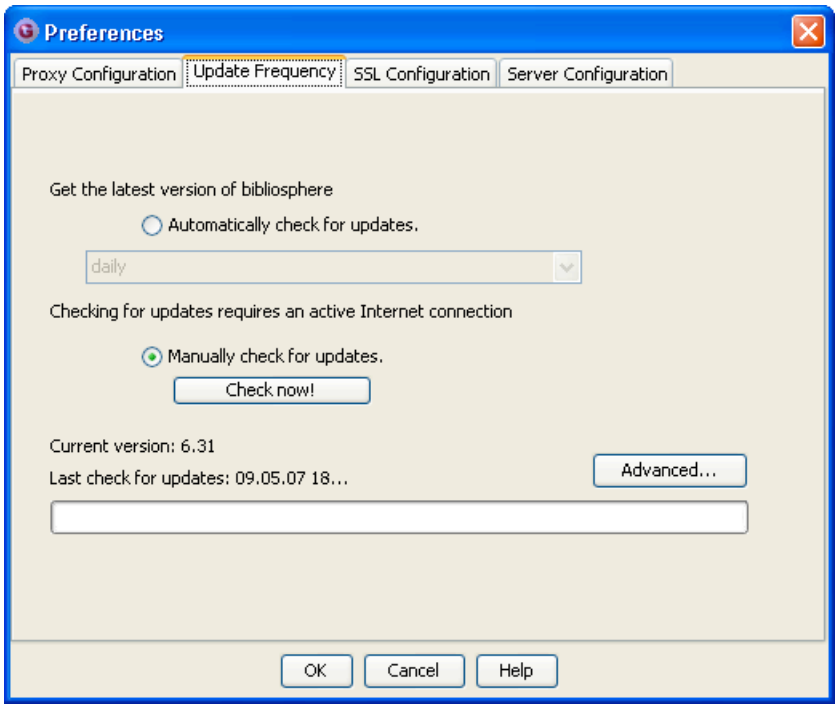
Click on the “Server Configuration”:



Check for Updates

Periodically Genomatix provides important BSPE updates. The Genomatix Update Service helps you to keep your application current.

Click on “Update Frequency” in the Configuration dialog.

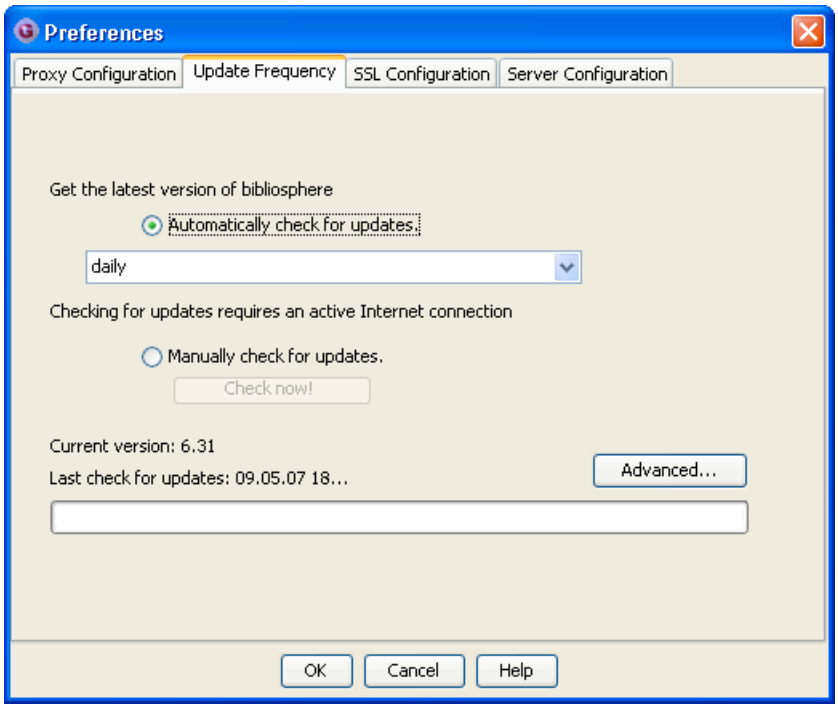


There are two modes for update: “Automatically check for updates” and “Manually check for updates”:

Turning on Automatic Update Notification

The Automatic Update Service checks for updates at regular intervals. Any time a product update becomes available, you receive a notification. Once you receive the notification, the Update Service guides you toward the download and installation of the updates you need. The Automatic Update Service is activated as follows:

Select **"automatically check for updates"** and choose your preferred update frequency (choices are **"daily", "weekly"** and **"monthly"**). Then press the **"ok"**-button.

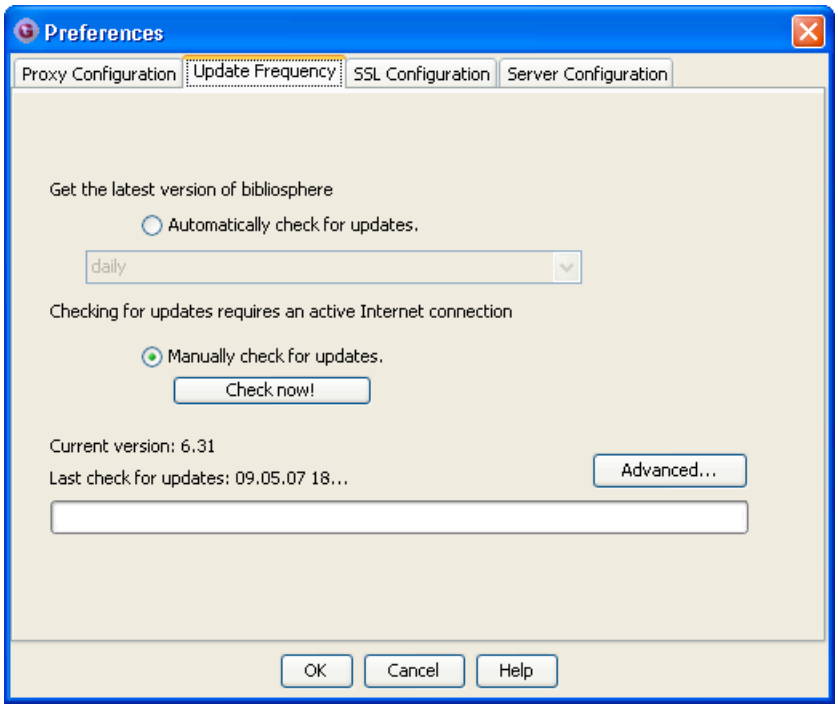


Updating your Application Manually

In some situations, you might want to update your application manually.

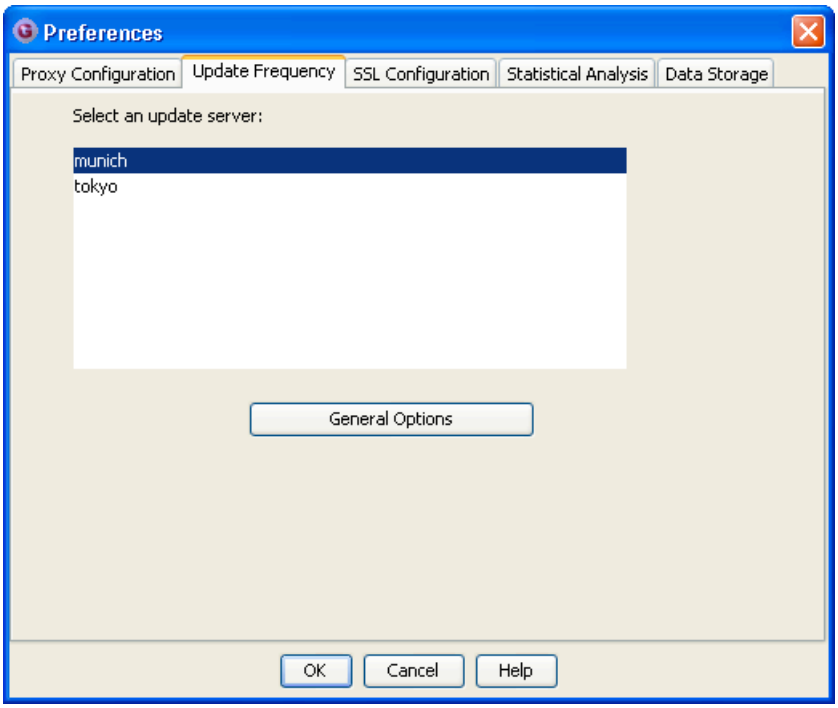
Select **"Manually check for updates".** This will activate the **"Check now"**-button.

Press the **"Check now"**-button. If an update is available the Update Service will guide you through the update process.



Selecting an Update Server

If update speed is slow, click the “Advanced...” button in the Update Frequency panel and select a different update server from the list. To go back to the main panel, click the “General Options” button.



Genomatix
understanding gene regulation

How to Prepare your Input Data

BSPE expects a list of genes (required) with a signal value assigned to every gene (optional). A gene list query only contains terms to identify genes. If signal values for the genes should be added for analysis, you have to create an excel file (see below).

Gene Identifiers

Gene identifiers can be entered as a list separated by white spaces, commas or semicolons.

The following gene identifiers are accepted by BSPE:

1. Gene symbols (e.g. icam3)
2. Gene description (e.g. mitogen-activated protein kinase 1)
3. Entrez gene identifier or GeneID (e.g. LOC5166 or 5166)
4. RefSeq ID (e.g. NM_02044)
5. GenBank oligo capped mRNAs (e.g. AK000539)
6. UniGene ID (e.g. Hs.202453)

BSPE allows using different gene identifiers in the same list.

If you do not remember the exact gene name or gene description you want to retrieve, you can use an asterisk (*) in your search term. The asterisk (*) represents a wildcard, meaning a placeholder for **zero or more** unknown characters.

Example: mapk* retrieves mapk1, mapk2, mapk12, mapk13, but also MAPK/ERK and “putative mapk”

Gene List

There are two different ways to enter a list of genes for analysis with BSPE:

- Gene list query which is directly entered in the BSPE user interface (without assigned value)
- Excel file which can be uploaded (with and without assigned value), e.g. as provided by Genomatix’s microarray analysis software, ChipInspector

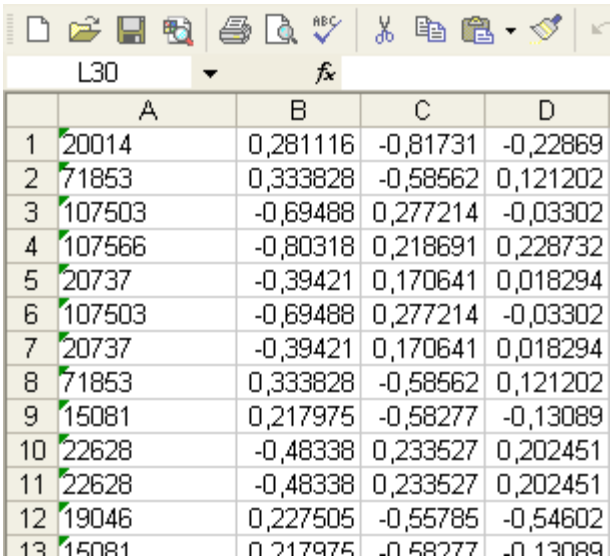
User Interface

You can enter a comma, semicolon, or space separated gene list directly in the interface.

© 2007 Genomatix Software GmbH

Excel Files

An Excel file requires one column with gene identifiers. Valid identifiers are: Entrez GeneIDs (e.g. 5166), Affymetrix probe set IDs (e.g. 202275_at), and Genomatix Transcript IDs (e.g. GXT_2740761). Optionally, you can place one or more columns containing any kind of numerical values after the identifier column (e.g. expression values of a microarray experiment).



L30

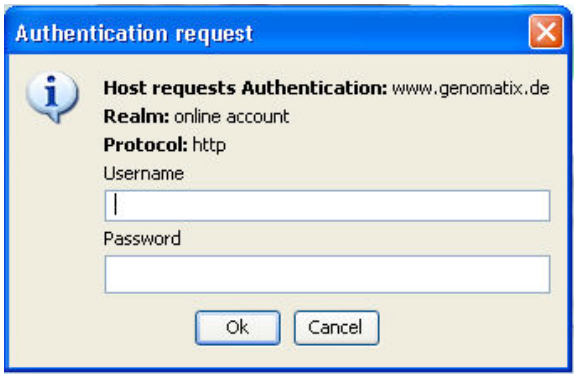
	A	B	C	D
1	20014	0,281116	-0,81731	-0,22869
2	71853	0,333828	-0,58562	0,121202
3	107503	-0,69488	0,277214	-0,03302
4	107566	-0,80318	0,218691	0,228732
5	20737	-0,39421	0,170641	0,018294
6	107503	-0,69488	0,277214	-0,03302
7	20737	-0,39421	0,170641	0,018294
8	71853	0,333828	-0,58562	0,121202
9	15081	0,217975	-0,58277	-0,13089
10	22628	-0,48338	0,233527	0,202451
11	22628	-0,48338	0,233527	0,202451
12	19046	0,227505	-0,55785	-0,54602
13	15081	0,217975	-0,58277	-0,13089

Starting an Analysis

Inputting your Data

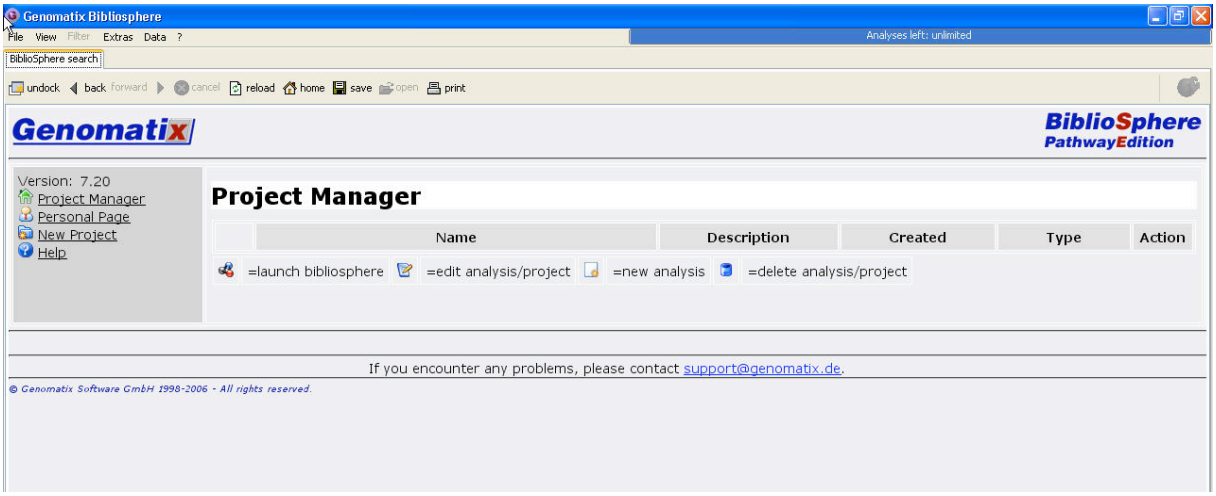
Sign In

Clicking the green arrow symbol on the BSPE start screen opens a dialog where you enter your username and password to connect to your BiblioSphere server.



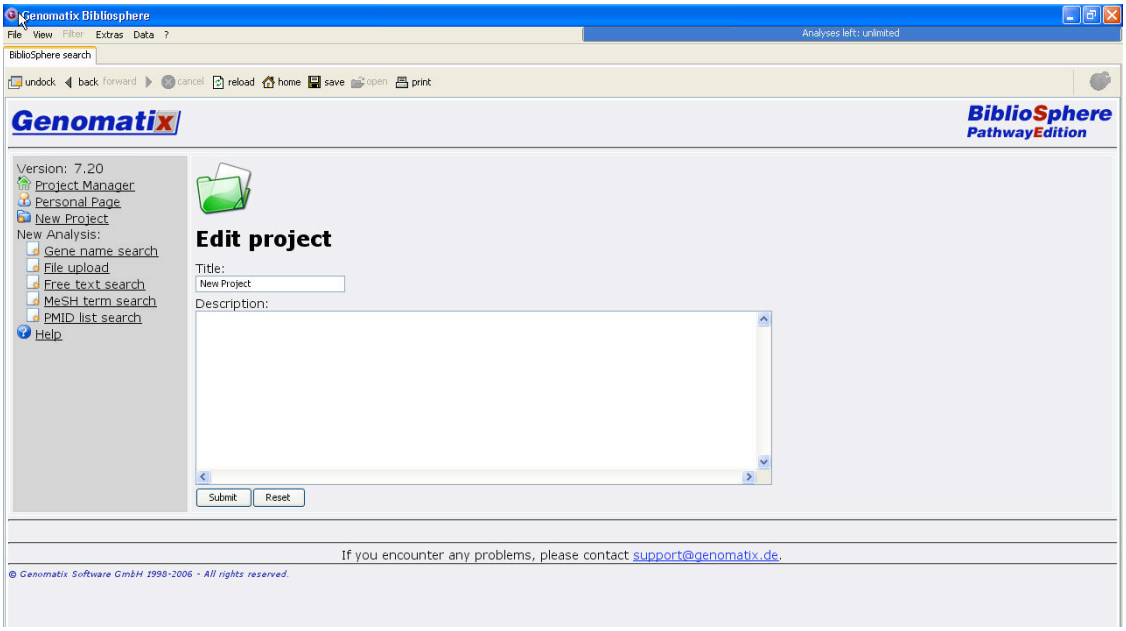
Project Management

You start off with your analyses in the Project Manager panel. Projects group your analyses for easier identification and retrieval of results.



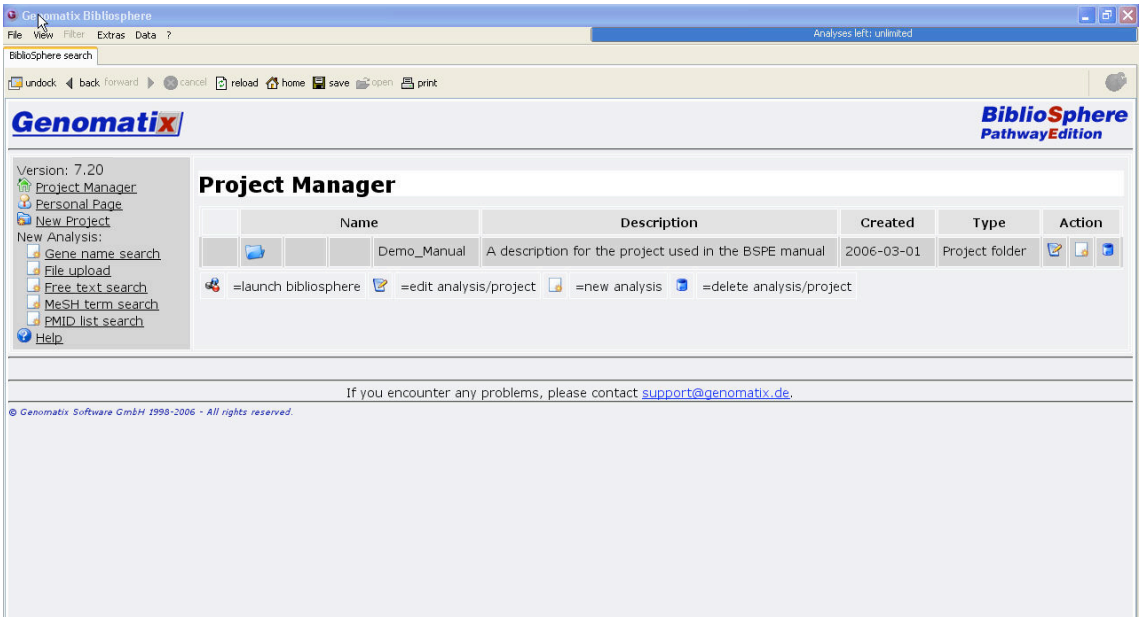
Create a New Project

Clicking “New Project” opens a panel where you can enter a name for your project and an optional description. The new project is added to the project list in the Project Manager panel.



Edit an Existing Project

You can edit an existing project, i.e. change its name or description, anytime. To open the editing panel, click the “edit analysis/project” symbol in the respective row. To delete a project, including all its associated analyses, click on the respective “delete analysis/project” symbol.



Create a New Analysis

You can add a new analysis to a project by clicking on the “new analysis” symbol in a project row – this will create a gene name search based analysis. Alternatively, you can select an analysis from the Project Manager menu.

There are five different ways to provide the data you want to analyze:

- Input a list of gene names or Gene IDs
- Upload an Excel file containing the gene names or IDs
- Enter free text search terms
- Enter a list of MeSH terms
- Input a list of PMIDs

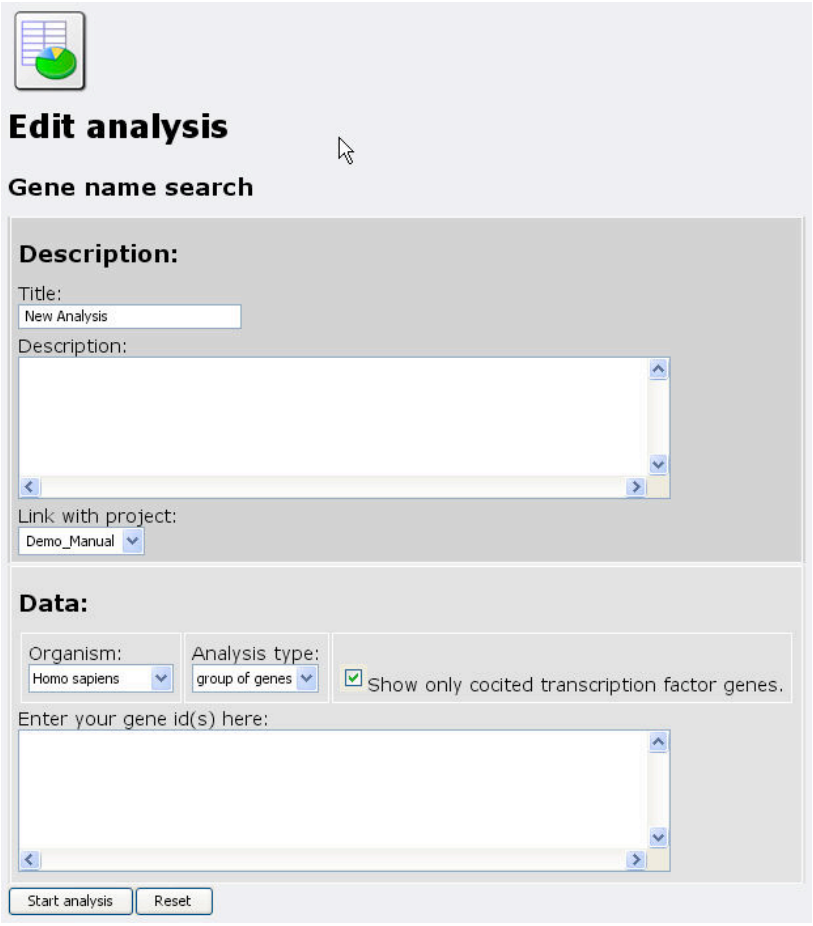
Common to all of them, you enter a title and an optional description for your analysis, and select the project you want to link it to. Also, you select the species in which you are going to search for genes connected to your input data. At present, available species are human, mouse, chicken, rat, zebrafish, chimpanzee, dog, cow, rhesus monkey and C. elegans.
Entering Gene Identifiers comprised of more than one word necessitates the use of quotation marks.

Input Data

BSPE searches for co-citations between your input genes. If you perform an analysis that requires you to specify gene identifiers as input (i.e., a gene name search or a file upload analysis), by default only those genes in your input set that are either co-cited with another gene in the input set or with any transcription factor are included in the resulting gene network.

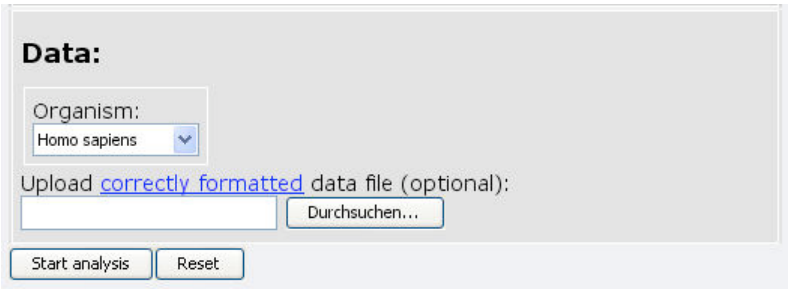
Gene name search

If you provide gene names, you can choose between two different types of analysis, single gene, and group of genes. A single gene analysis will retrieve a Single Gene Centred BiblioSphere (SGBS) for each of your input genes from the database, while a group centred analysis additionally will generate a Cluster Centred BiblioSphere (CCBS) based on all your input genes. You can switch between cluster and gene centred views in an analysis of the latter type.
If you keep the “Show only co-cited transcription factor genes” option checked, which is the default setting, the input genes that are either co-cited with another input gene or with any transcription factor (which is not necessarily among the input genes), as well as the co-cited transcription factors, will be included in the CCBS. Other genes that are co-cited with an input gene, but do not code for transcription factors, will not be included, nor will input genes for which there is no co-citation with another input gene or with a transcription factor.
Deselecting the “Show only co-cited transcription factor genes” option has two effects: Firstly, any gene that is co-cited with one of your input genes will be included in the network, regardless of its coding for a transcription factor, and secondly, any co-citation of an input gene is sufficient qualification for inclusion of that gene in the CCBS. The SGBS remain unaffected by this setting.



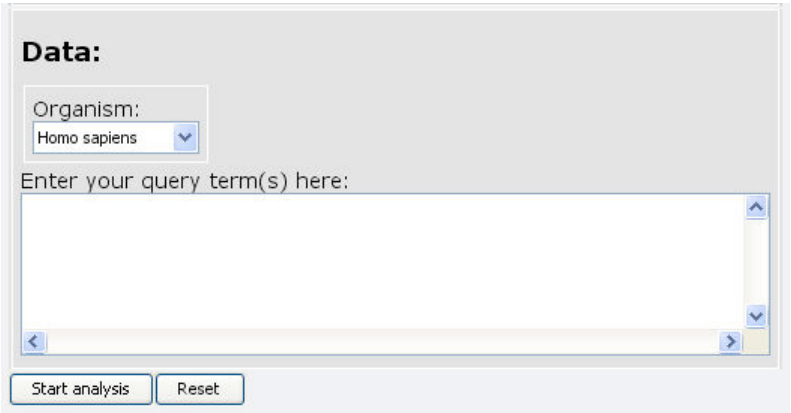
File upload

You can let BSPE read a gene list from an Excel file, whose path and name you can enter here. As to the accepted format, see "How to prepare your input data – Excel Files". The analysis includes both CCBS and SGBS based on the input genes.



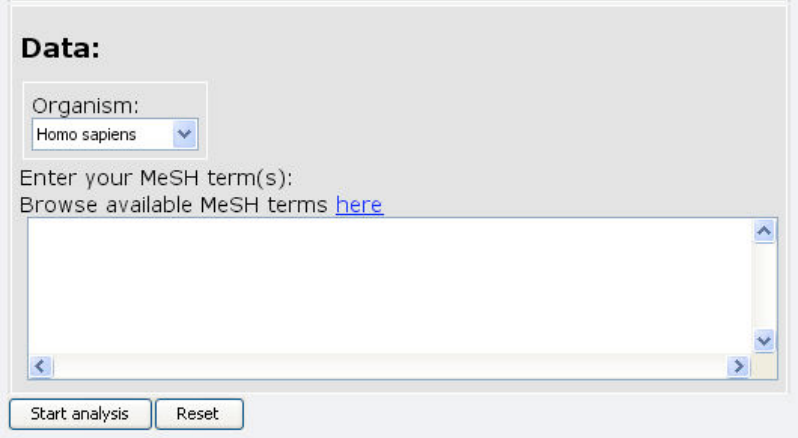
Free text search

You can enter one or more search terms that will be combined using the OR operator by default. However, you may also explicitly specify the logical operators. Accepted operators are: AND, OR, NOT, + , and -; use of parentheses is possible. A free text search based analysis creates a CCBS comprised of the genes mentioned in the articles' abstracts that were found using the search terms.



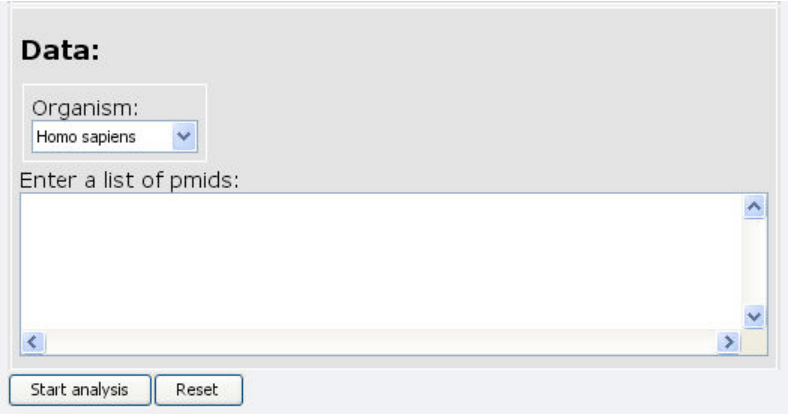
MeSH term search

You can search by one or more valid MeSH terms; a link for browsing available MeSH terms is provided. As to use of Boolean operators, the same rules as in the free text search apply. The genes displayed in the resulting CCBS will be those that appear in the articles found with the selected MeSH search terms.



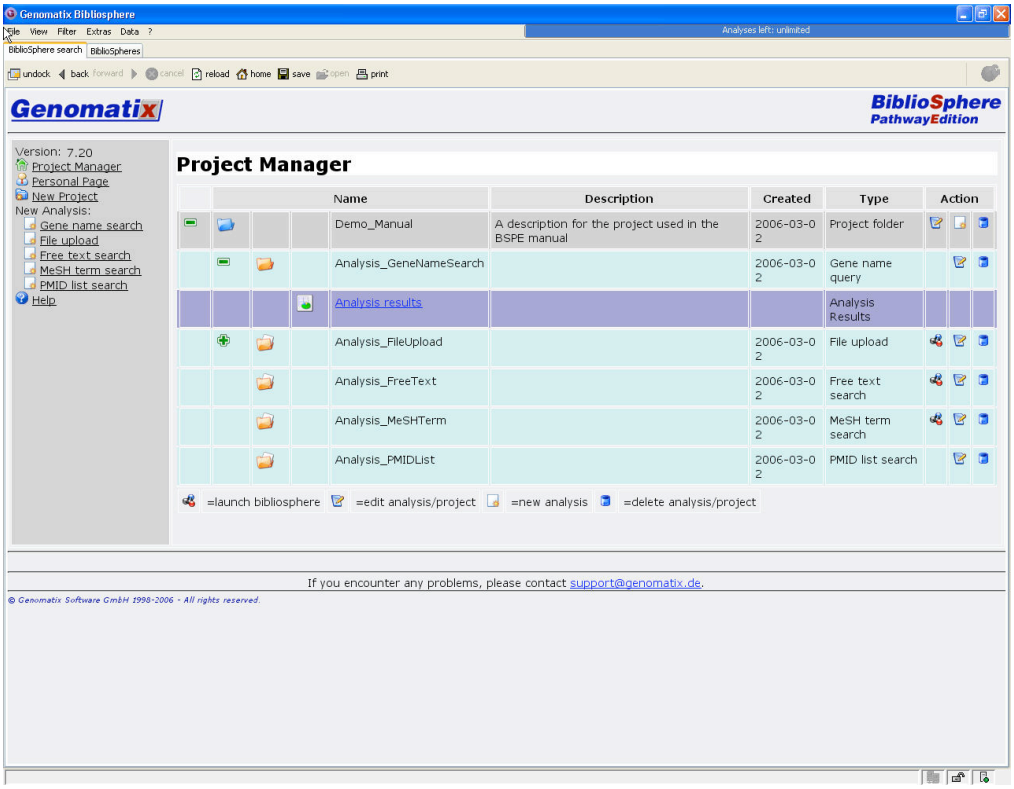
PMID list search

Here you can enter a list of PubMed IDs to search based on the genes covered in the pertaining articles' abstracts. The genes that appear in the articles with the selected PMIDs will be displayed in a CCBS.



Accessing your Analyses

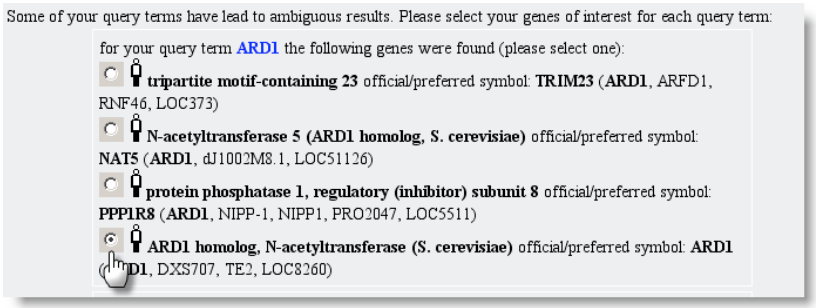
Any new analysis will appear in the Project Manager panel under the project it has been associated to. If a CCBS was generated during the analysis, you will be able to retrieve it anytime by clicking the "launch bibliosphere" icon in the respective row. If a list of SGBSs exists, it is accessible via the "Analysis results" hyperlink in the row below the analysis.
If you want to update or delete an analysis, click the according symbol in the appropriate row.



Running your Analysis

Ambiguities

It is possible that BSPE returns a list of proposed genes if the input identifier is ambiguous. In this case, a list similar to the following one appears, asking you to enter the correct correspondence. The official/preferred symbol is the default choice.

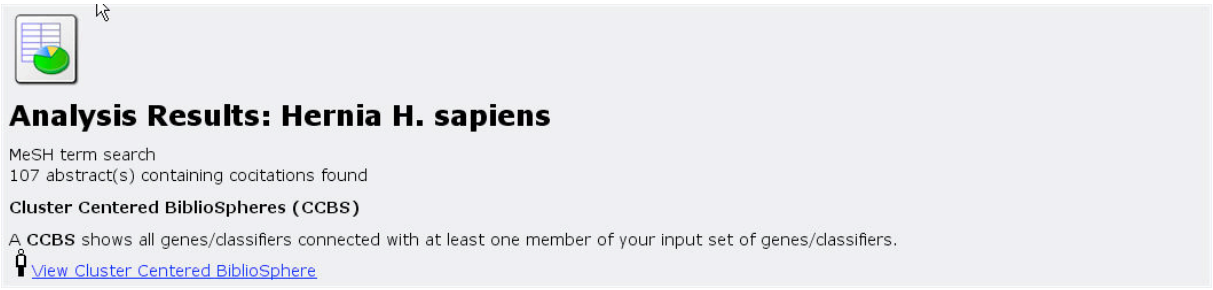

Some of your query terms have lead to ambiguous results. Please select your genes of interest for each query term:
for your query term ARD1 the following genes were found (please select one):
tripartite motif-containing 23 official/preferred symbol: TRIM23 (ARD1, ARFD1, RNF46, LOC373)
N-acetyltransferase 5 (ARD1 homolog, S. cerevisiae) official/preferred symbol: NAT5 (ARD1, dJ1002M8.1, LOC51126)
protein phosphatase 1, regulatory (inhibitor) subunit 8 official/preferred symbol: PPP1R8 (ARD1, NIPP-1, NIPP1, PRO2047, LOC5511)
ARD1 homolog, N-acetyltransferase (S. cerevisiae) official/preferred symbol: ARD1
(ARD1, DXS707, TE2, LOC8260)

This can happen if ambiguous gene descriptions/symbols are used, or if one name is used for different genes (homonym). It is always recommended to use the unambiguous Locus ID for the input genes.

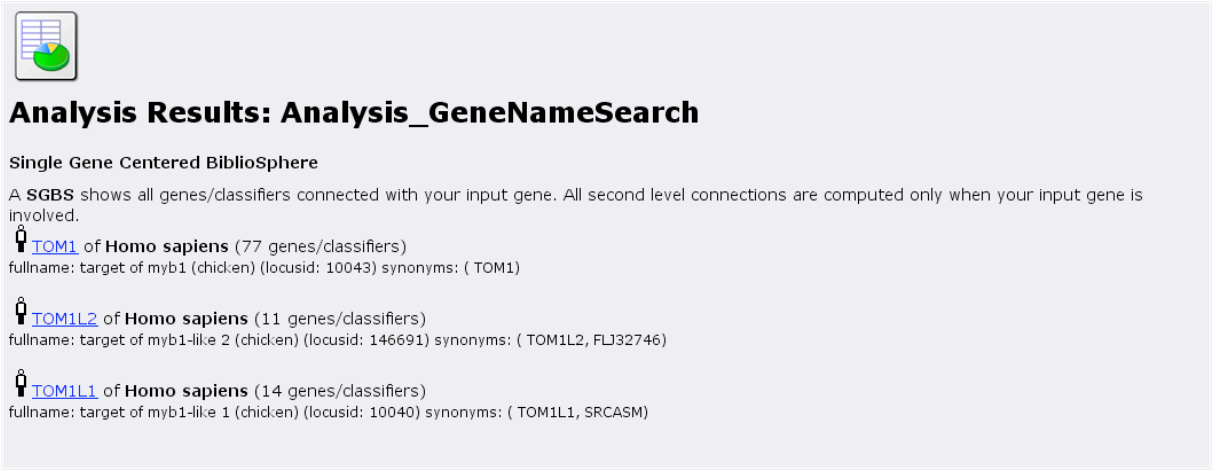
Hit List

Depending on the type of analysis performed, BSPE displays either information on the generated CCBS, or an SGBS list, or both.

Example CCBS info:


Analysis Results: Hernia H. sapiens
MeSH term search
107 abstract(s) containing cocitations found
Cluster Centered BiblioSpheres (CCBS)
A CCBS shows all genes/classifiers connected with at least one member of your input set of genes/classifiers.
View Cluster Centered BiblioSphere

Example SGBS list:

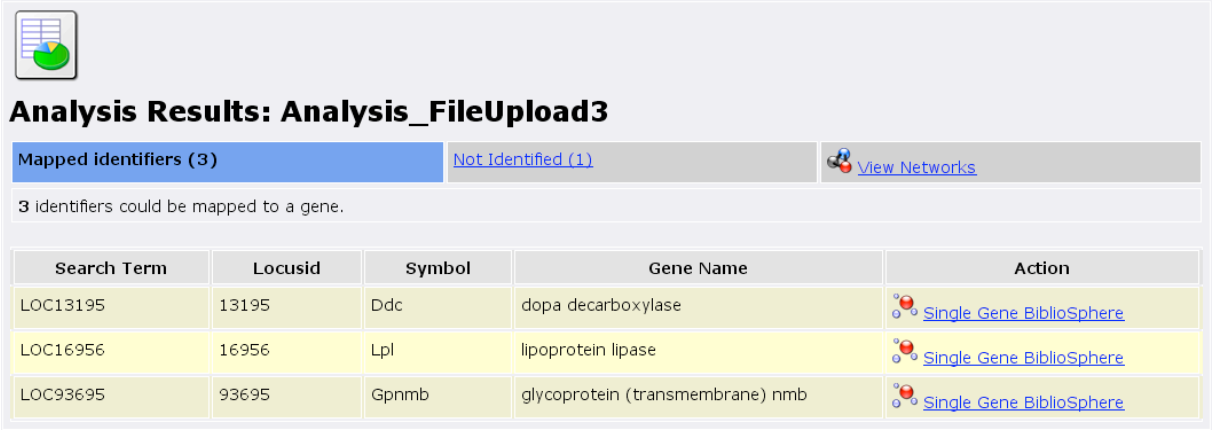

Analysis Results: Analysis_GeneNameSearch
Single Gene Centered BiblioSphere
A SGBS shows all genes/classifiers connected with your input gene. All second level connections are computed only when your input gene is involved.
TOM1 of Homo sapiens (77 genes/classifiers)
fullname: target of myb1 (chicken) (locusid: 10043) synonyms: (TOM1)
TOM1L2 of Homo sapiens (11 genes/classifiers)
fullname: target of myb1-like 2 (chicken) (locusid: 146691) synonyms: (TOM1L2, FLJ32746)
TOM1L1 of Homo sapiens (14 genes/classifiers)
fullname: target of myb1-like 1 (chicken) (locusid: 10040) synonyms: (TOM1L1, SRCASM)

Both a CCBS and an SGBS list are available if the gene identifiers were provided by file upload or if “group of genes” was selected in a gene name based search. If any of the input gene identifiers were not recognized by BiblioSphere, they are listed in an extra table. You can switch between these different views.

Example CCBS view:

Analysis Results: Analysis_FileUpload3
Mapped identifiers (3)
Not Identified (1)
View Networks
3 identifiers could be mapped to a gene.
Cluster Centered BiblioSpheres (CCBS)
A CCBS shows all genes/classifiers connected with at least one member of your input set of genes/classifiers.
View Cluster Centered BiblioSphere

Example SGBS view:



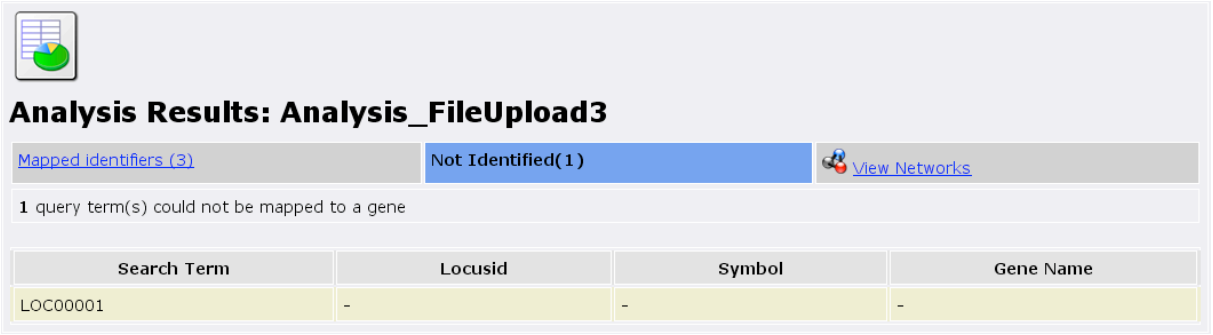
Analysis Results: Analysis_FileUpload3

Mapped identifiers (3) | Not Identified (1) | View Networks

3 identifiers could be mapped to a gene.

Search Term	Locusid	Symbol	Gene Name	Action
LOC13195	13195	Ddc	dopa decarboxylase	Single Gene BiblioSphere
LOC16956	16956	Lpl	lipoprotein lipase	Single Gene BiblioSphere
LOC93695	93695	Gpnmb	glycoprotein (transmembrane) nmb	Single Gene BiblioSphere

Example Unidentified Search Term view:



Analysis Results: Analysis_FileUpload3

Mapped identifiers (3) | Not Identified(1) | View Networks

1 query term(s) could not be mapped to a gene

Search Term	Locusid	Symbol	Gene Name
LOC00001	-	-	-

Network View

Pathway View

Overview

A BiblioSphere represents your input gene's bibliographical environment. It contains your input gene(s), genes co-cited with the input genes, and various information pertaining to the genes, the relationships between them, and their literature context. Various filter settings allow you to restrict the view to elements of your interest.

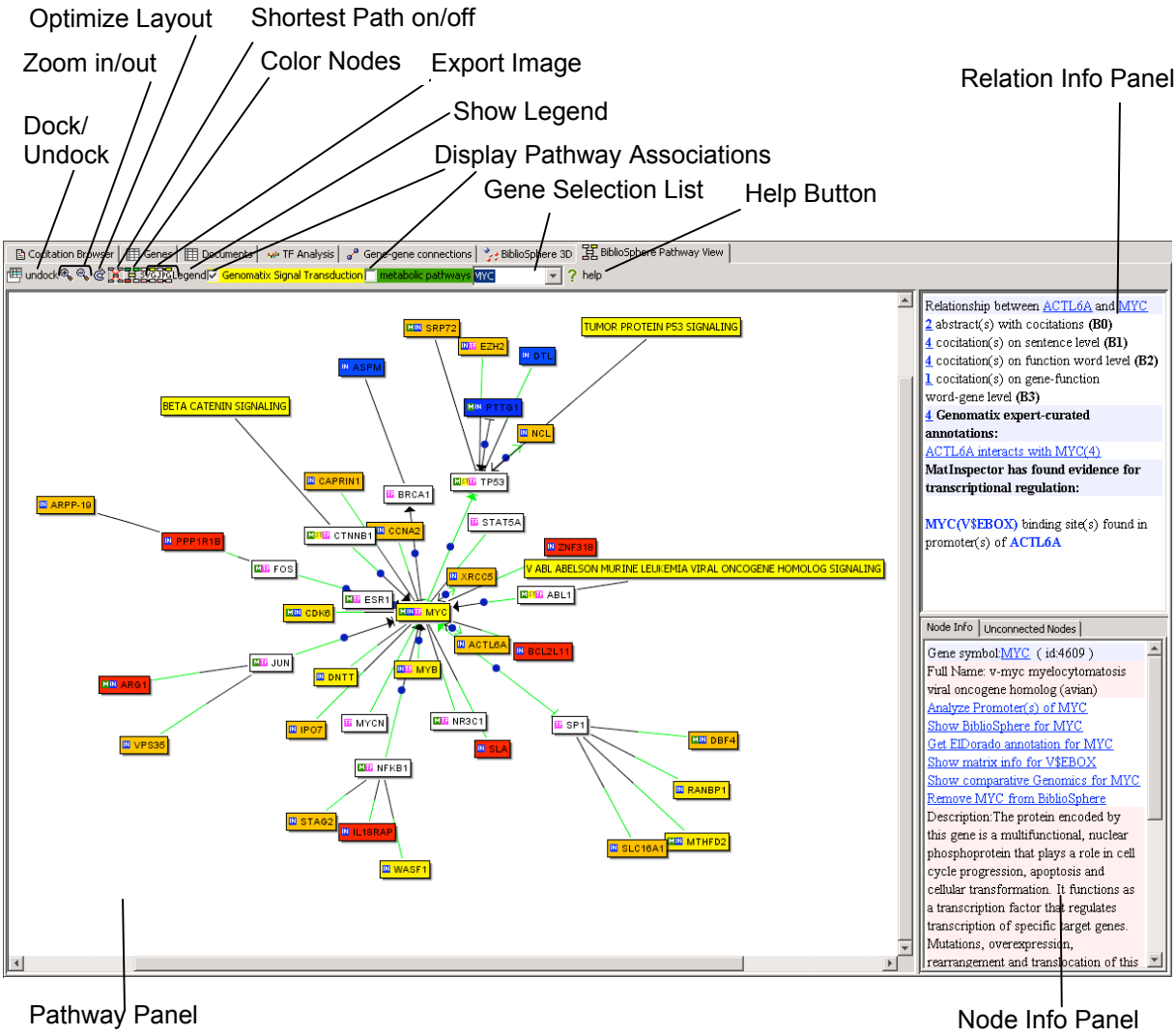
You open a BiblioSphere view by clicking on the respective link in the BiblioSphere search view. Several BiblioSpheres can be open in parallel. You can switch between them by clicking the appropriate tab. SGBS tabs are labelled with the input gene ID, CCBS tabs with the name of the analysis they were created for. An icon denotes the species selected in the analysis.

The workspace area of a BiblioSphere view consist of three panels, a filter panel on the left hand side, a protocol panel at the bottom, and the main panel, which occupies the rest of the available space. The main panel itself contains several different views on the data, organized in tabbed panes.

The BiblioSphere Pathway View pane is displayed in the foreground when you open a BiblioSphere. It shows a graphical network representation of the BiblioSphere in the Pathway Panel. Genes are displayed as network nodes, whereas the relationships between them make up the edges. If you click on a relationship, information on it is shown in the upper right hand corner Relation Info panel. Below that, details on the gene that currently has the focus in the network view are displayed in the Node Info tab of the Node Info Panel. In the Unconnected Nodes tab, all input genes that could not be integrated into the network are listed.

Toolbar Controls and Actions:

	Zoom In:	Zoom in to the Pathway View.
	Zoom Out	Zoom out.
	Optimize Layout	Recalculate the layout depending on the gene in focus.
	Shortest Path	Shortest path for the gene in focus ON/OFF.
	Color Scheme Chooser	Launch the Color Scheme Chooser (only enabled when expression/ranking data is available).
	Save SVG	Save an image of the network as Scalable Vector Graphic (SVG).
	Save JPG	Save a bitmap image of the network in JPEG format.
Legend	Display Legend	Display a legend for gene-gene connections.
Check box	Genomatix Signal Transduction	Display signal transduction pathway associations of visible genes.
Check box	Metabolic Pathways	Display BioCyc metabolic pathway associations of visible genes.
Pull down menu	Gene Selection List:	Sorted list of network genes. Focus gene by selecting it from the list.



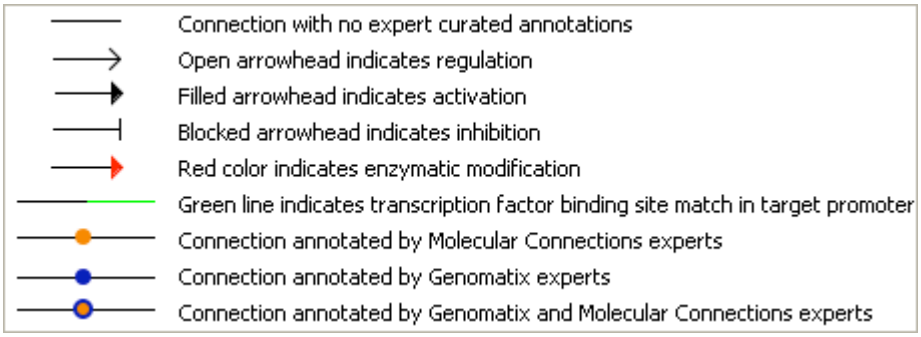
Display symbols for input genes are highlighted blue, currently selected genes dark blue.
Different symbols are used to mark special properties:

TF	Gene product is a transcription factor
ST	Gene product is part of a Genomatix signal transduction pathway
IN	Input gene (only in CCBS)
M	Gene product is part of a metabolic pathway
U	User annotated gene

Position the mouse pointer over a gene symbol to display the gene's full name and Gene ID.

Connection Modes

Functional relationships between co-cited genes are visualized by the connection lines between the genes:


Connection with no expert curated annotations
Open arrowhead indicates regulation
Filled arrowhead indicates activation
Blocked arrowhead indicates inhibition
Red color indicates enzymatic modification
Green line indicates transcription factor binding site match in target promoter
Connection annotated by Molecular Connections experts
Connection annotated by Genomatix experts
Connection annotated by Genomatix and Molecular Connections experts

Arrowheads at the ends of a connecting line symbolize the type of functional relationship between the connected genes.

If a gene that codes for a transcription factor is connected to a gene that is known to contain a binding site for this transcription factor in its promoter, the connecting line is colored green over half of its length near the gene containing the binding site.

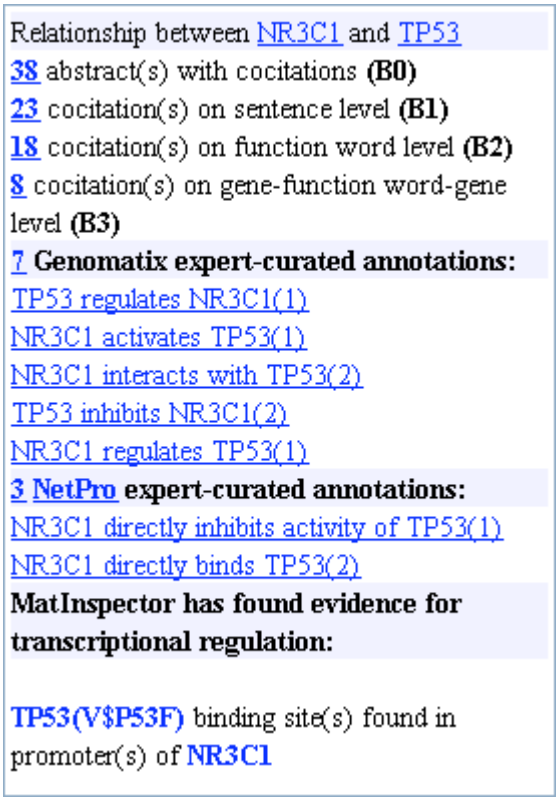
Hand-annotated gene-gene relationships are indicated by a circle in the centre of the connection line.

Relation Info Panel

The Relation Info Panel shows information on the currently selected connection between two genes, particularly on the numbers of co-citations of these genes on different levels. Available levels are: Abstract, Sentence, Function Word, Gene-Function Word-Gene (GFG), and Expert. For detailed information on co-citation levels, see the chapter "Co-citation Filter". The information provided in detail is:

- Genes that are linked by the currently selected connection. Clicking on a gene symbol hyperlink opens a browser window with detailed gene information.
- The number of abstracts containing co-citations of the connected genes. Clicking this number opens the Cocitation Browser, which provides links to the abstracts, and which will display every sentence in the abstracts that cites at least one of the co-cited genes.
- For each co-citation level above Abstract, the number of co-citing sentences that conform to that level. Clicking on a number will open the Cocitation Browser, displaying the relevant sentences.
- Expert-curated annotations, with links that will display the pertaining sentences in the Cocitation Browser.
- If one of the connected genes codes for a transcription factor, matching binding sites in the promoter of the other gene, including the binding site's matrix family name.

Example Relation Info Panel contents:


Relationship between NR3C1 and TP53
38 abstract(s) with cocitations (B0)
23 cocitation(s) on sentence level (B1)
18 cocitation(s) on function word level (B2)
8 cocitation(s) on gene-function word-gene level (B3)
7 Genomatix expert-curated annotations:
TP53 regulates NR3C1(1)
NR3C1 activates TP53(1)
NR3C1 interacts with TP53(2)
TP53 inhibits NR3C1(2)
NR3C1 regulates TP53(1)
3 NetPro expert-curated annotations:
NR3C1 directly inhibits activity of TP53(1)
NR3C1 directly binds TP53(2)
MatInspector has found evidence for transcriptional regulation:
TP53(V$P53F) binding site(s) found in promoter(s) of NR3C1

Node Info Panel

In the BiblioSphere Pathway View, the information displayed in the Node Info Panel is distributed to two tabs, whereas in the BiblioSphere 3D View the Node Info Panel displays only Gene node (see below) information and is therefore not tabbed.

Node Info Tab

The Node Info tab shows information for the currently selected node in the BiblioSphere graph. Nodes can represent either a gene or a pathway annotation for a gene.

Gene node

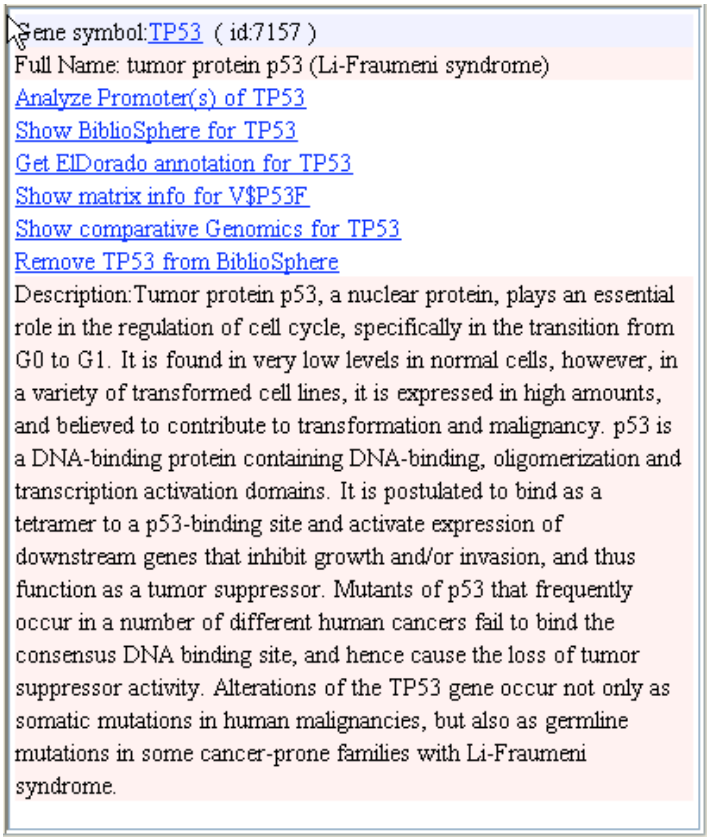
The gene's locus ID, full name, organism, and description are displayed. Clicking on the gene symbol hyperlink opens a browser window with detailed gene information.

Moreover, links to other components of the Genomatix Suite are provided that offer further analysis of the gene:

- Promoter analysis with MatInspector
- Single Gene Centred BiblioSphere
- ElDorado annotation
- Comparative genomics with ElDorado
- TF binding site matrix (for transcription factors)

Another link allows you to remove the gene from the BiblioSphere

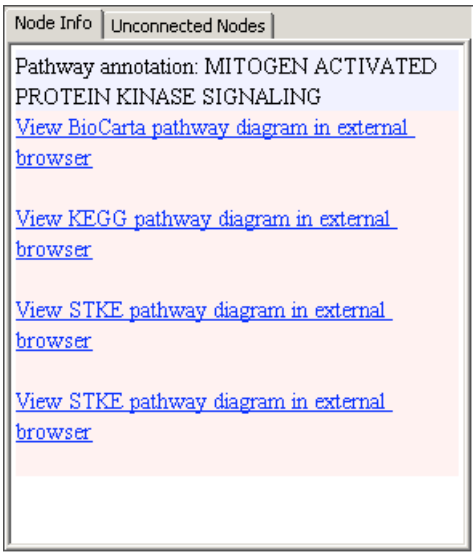
Example gene node information:


Gene symbol:TP53 (id:7157)
Full Name: tumor protein p53 (Li-Fraumeni syndrome)
Analyze Promoter(s) of TP53
Show BiblioSphere for TP53
Get ElDorado annotation for TP53
Show matrix info for V$P53F
Show comparative Genomics for TP53
Remove TP53 from BiblioSphere
Description:Tumor protein p53, a nuclear protein, plays an essential role in the regulation of cell cycle, specifically in the transition from G0 to G1. It is found in very low levels in normal cells, however, in a variety of transformed cell lines, it is expressed in high amounts, and believed to contribute to transformation and malignancy. p53 is a DNA-binding protein containing DNA-binding, oligomerization and transcription activation domains. It is postulated to bind as a tetramer to a p53-binding site and activate expression of downstream genes that inhibit growth and/or invasion, and thus function as a tumor suppressor. Mutants of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA binding site, and hence cause the loss of tumor suppressor activity. Alterations of the TP53 gene occur not only as somatic mutations in human malignancies, but also as germline mutations in some cancer-prone families with Li-Fraumeni syndrome.

Pathway annotation node

For pathway annotations, the annotation is displayed.

Genomatix signal transduction pathway annotations can have outbound links to BioCarta, KEGG, and/or STKE pathway diagrams. Clicking a pathway link opens the according pathway graph on the content provider's web page in a browser window.


Node Info
Unconnected Nodes
Pathway annotation: MITOGEN ACTIVATED PROTEIN KINASE SIGNALING
View BioCarta pathway diagram in external browser
View KEGG pathway diagram in external browser
View STKE pathway diagram in external browser
View STKE pathway diagram in external browser

Unconnected Nodes Tab

The Unconnected Nodes tab displays all input genes that are not part of the generated network.

Docking and Undocking

You can move the Pathway View to a dedicated window by clicking the “undock” button. Clicking the “dock” button in that window or closing the window will re-dock the Pathway View.

Zoom

You can zoom in and out of the displayed network by clicking repeatedly on the zoom buttons in the toolbar.

Shortest Path

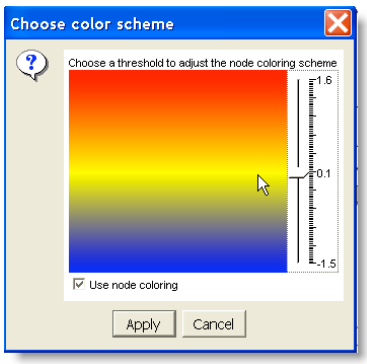
If the Shortest Path option is activated, display of edges is restricted to those that constitute the shortest path from the node that was double clicked last to all other displayed genes. Otherwise, all direct connections are displayed. Double clicking another node recalculates the shortest path with that node as the starting point and redisplays the graph. To optimize the layout after recalculating, click the Layout Optimization button.

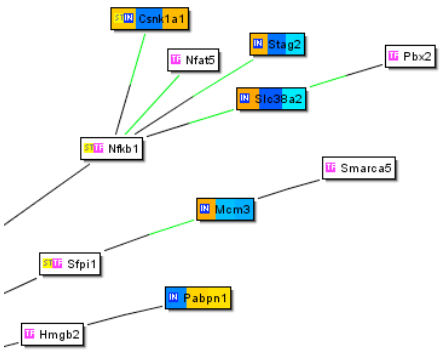
Layout Optimization

Clicking the Layout Optimization button redraws the network graph so that overlapping of connections and overall connection length are minimized, and relative connection lengths reflect the strengths of the connections optimally.

Color Scheme Chooser

The Color Scheme chooser can be used to activate and adjust the color coding for over and under expression of the genes in the BiblioSphere, if expression level information has been provided in the Excel file uploaded for analysis. Nodes for overexpressed genes are colored red, those for underexpressed genes blue. If expression values from a multi-class analysis were provided, the gene nodes will appear striped, displaying one color per experimental class. Moving the slider will change the color threshold for over/underrepresented genes.





Export Networks

Clicking the Save SVG or Save JPEG button lets you save the network graph in the respective format. In the dialog, please enter the file name including the extension.

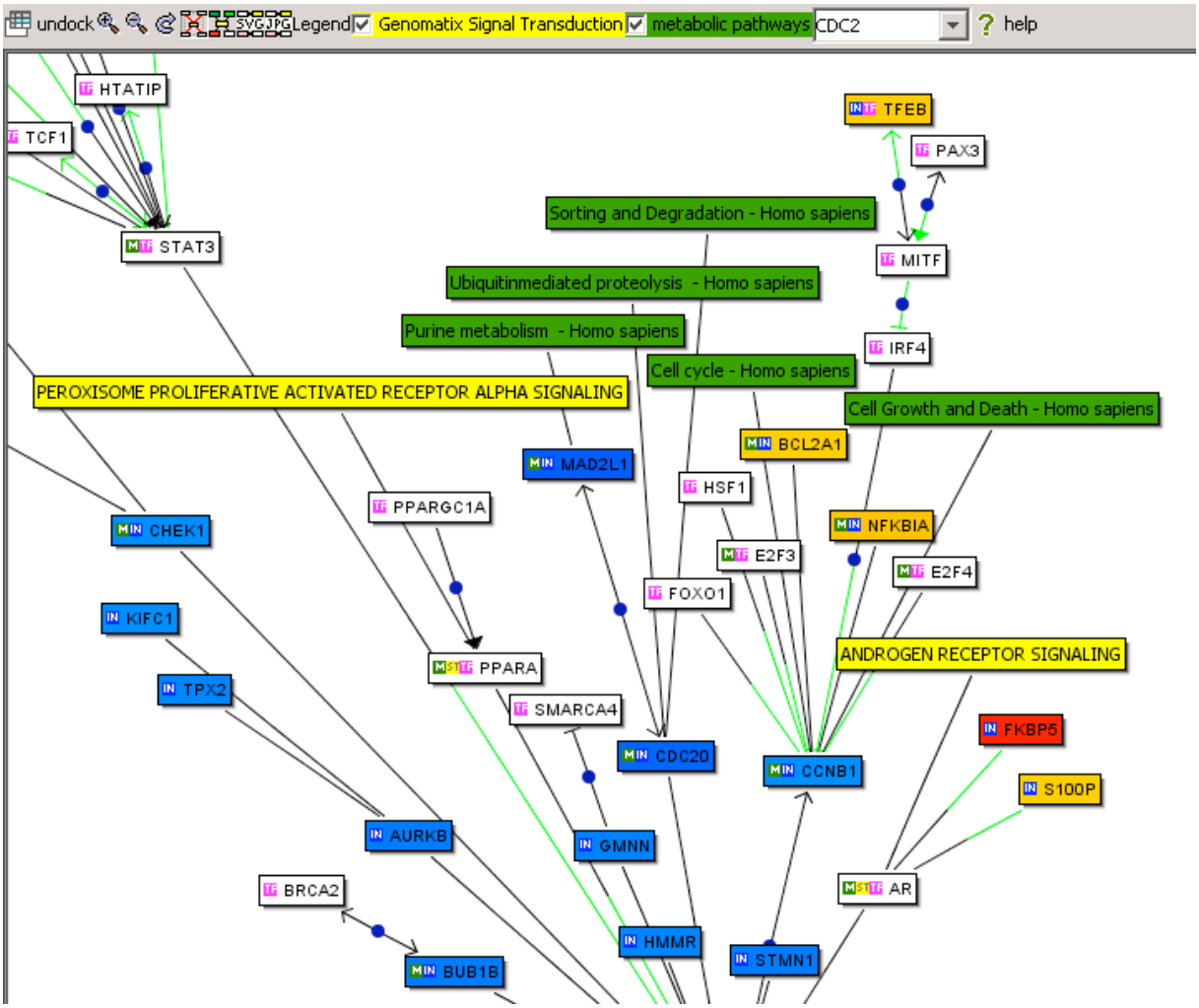
Metabolic & Signal Transduction Pathways

BSPE offers pathway annotations for the genes in the BiblioSphere. You can select to display the following annotations:

- Genomatix Signal Transduction pathway database
- BioCyc Metabolic Pathways database

Each option is available if at least one of the genes in the selected BiblioSphere has a pathway annotation of the respective type. It is not necessary that the annotated gene be displayed with the current filter settings, i.e., if you restricted the view on the genes, you might not see any pathway annotation of a certain type, even though you enabled their display.

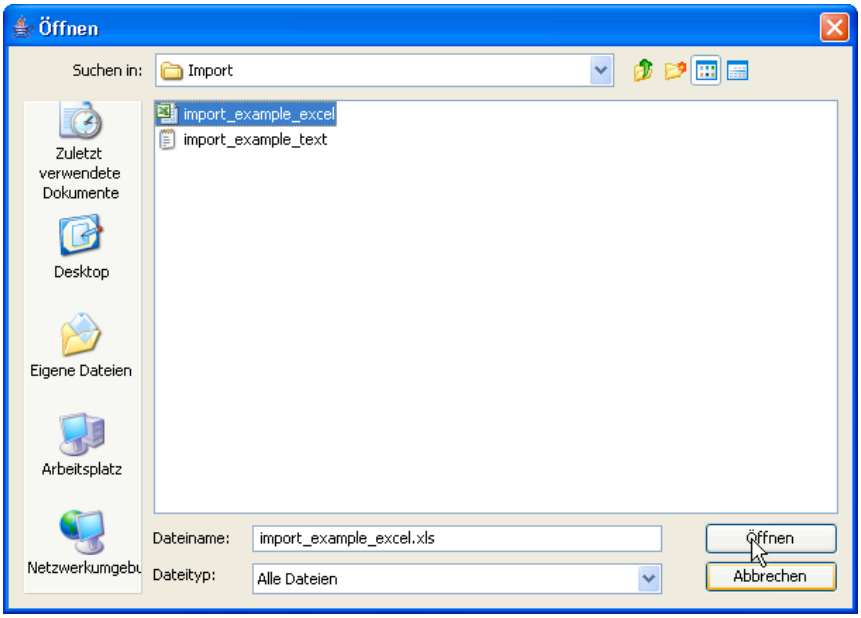
Pathway annotations appear in the graph as nodes that are linked to the annotated gene:



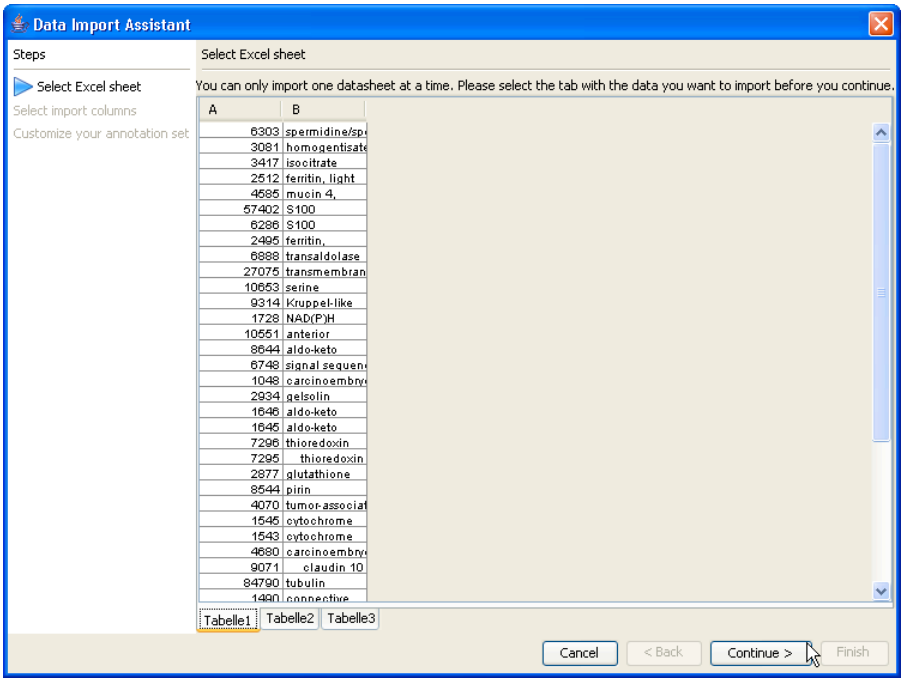
Clicking an annotation node will display information in the Node Info tab.

Importing your Own Annotations

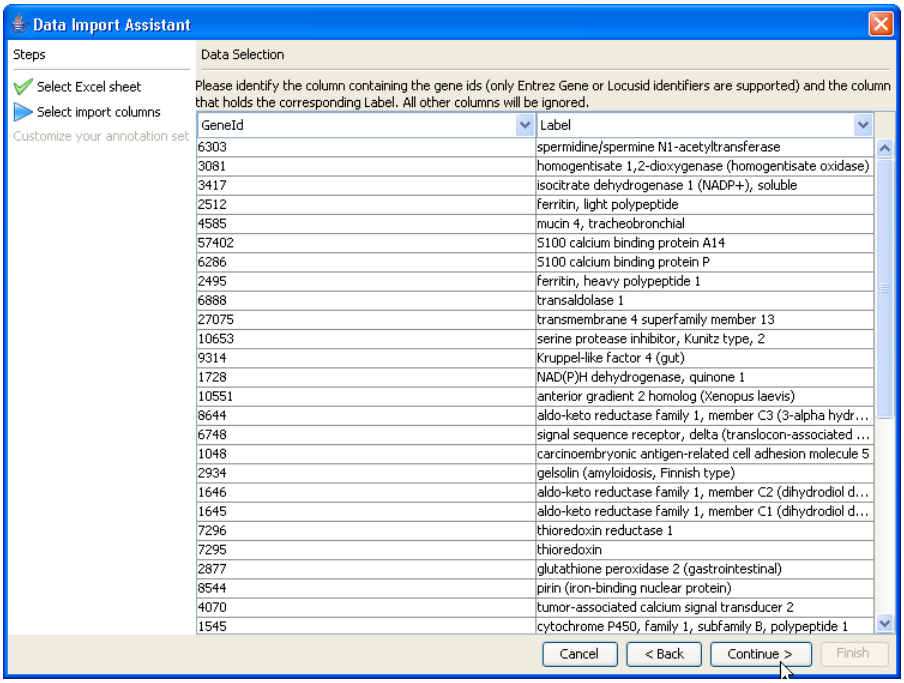
You can add your own sets of annotations to BSPE; they will appear in the Pathway view. To this end, select “Data – Import Annotations” from the main menu; a file dialog will be displayed:



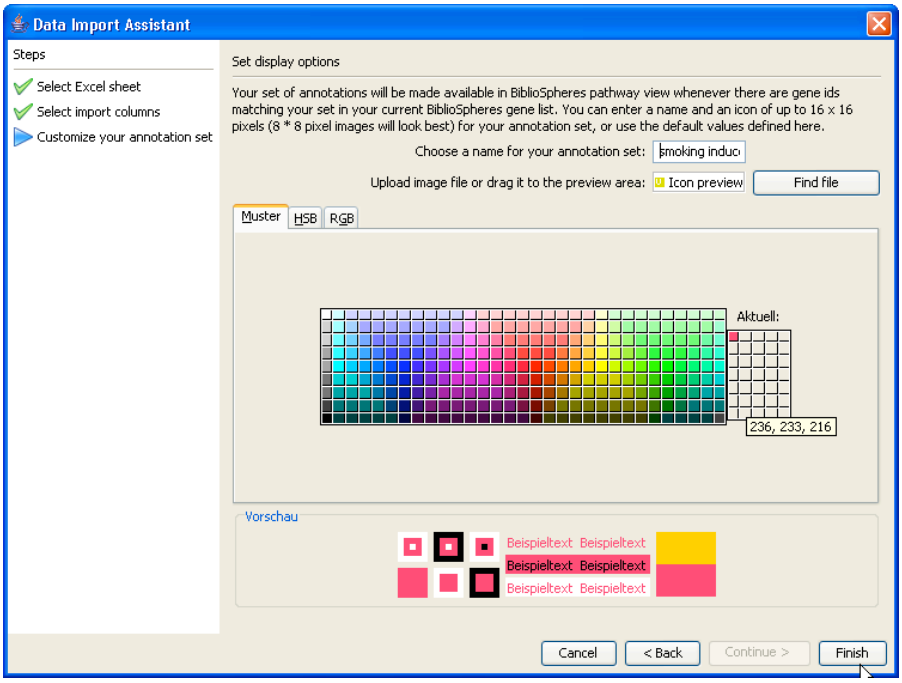
Your annotation file should be either an Excel file or a plain text file containing Gene IDs and the labels for each gene in the same row. Selecting a file will launch the Data Import Assistant. Depending on the format of your file, the assistant will guide you through the import process. If the selected file is an Excel file, you select the data sheet you would like to import:



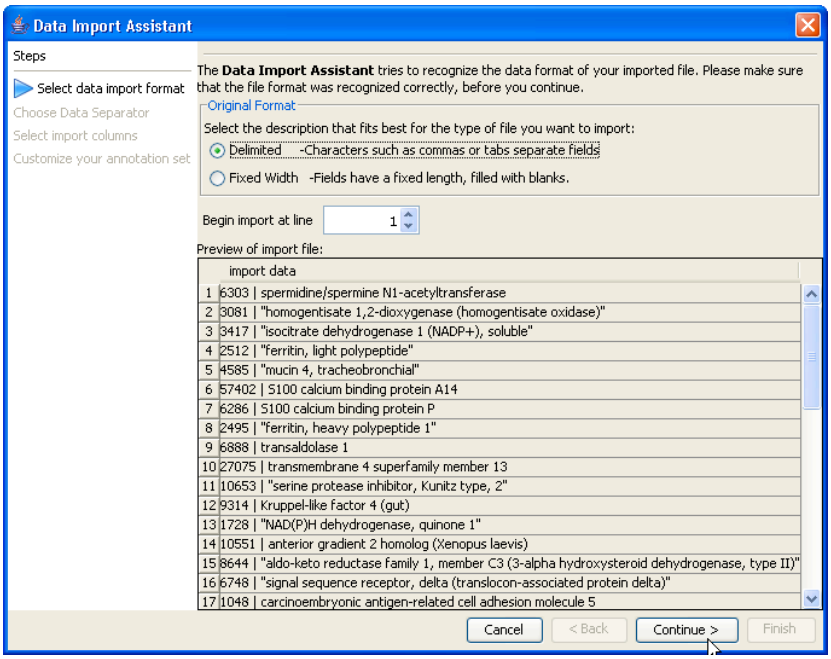
In the next step you can identify the column containing the Gene IDs, and the column holding the associated annotations by selecting the appropriate label in the column header. All other columns will be ignored:


Data Import Assistant
Steps
Select Excel sheet
Select import columns
Customize your annotation set
Data Selection
Please identify the column containing the gene ids (only Entrez Gene or Locusid identifiers are supported) and the column that holds the corresponding Label. All other columns will be ignored.
GeneId
Label
6303 spermidine/spermine N1-acetyltransferase
3081 homogentisate 1,2-dioxygenase (homogentisate oxidase)
3417 isocitrate dehydrogenase 1 (NADP+), soluble
2512 ferritin, light polypeptide
4585 mucin 4, tracheobronchial
57402 S100 calcium binding protein A14
6286 S100 calcium binding protein P
2495 ferritin, heavy polypeptide 1
6888 transaldolase 1
27075 transmembrane 4 superfamily member 13
10653 serine protease inhibitor, Kunitz type, 2
9314 Kruppel-like factor 4 (gut)
1728 NAD(P)H dehydrogenase, quinone 1
10551 anterior gradient 2 homolog (Xenopus laevis)
8644 aldo-keto reductase family 1, member C3 (3-alpha hydr...
6748 signal sequence receptor, delta (translocon-associated ...
1048 carcinoembryonic antigen-related cell adhesion molecule 5
2934 gelsolin (amyloidosis, Finnish type)
1646 aldo-keto reductase family 1, member C2 (dihydrodiol d...
1645 aldo-keto reductase family 1, member C1 (dihydrodiol d...
7296 thioredoxin reductase 1
7295 thioredoxin
2877 glutathione peroxidase 2 (gastrointestinal)
8544 pirin (iron-binding nuclear protein)
4070 tumor-associated calcium signal transducer 2
1545 cytochrome P450, family 1, subfamily B, polypeptide 1
Cancel
< Back
Continue >
Finish

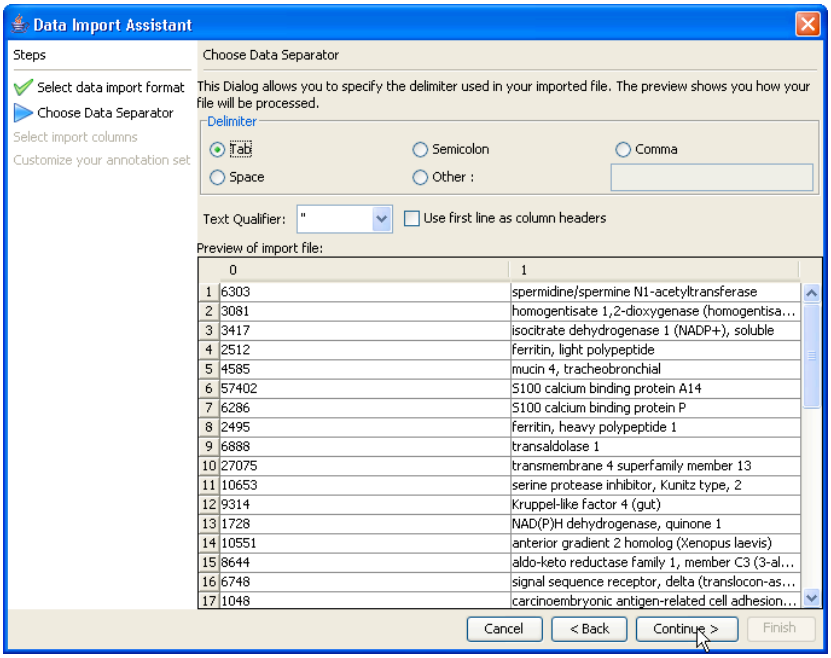
The last step allows you to name your annotation set and choose an icon that identifies genes with an annotation:


Data Import Assistant
Steps
Select Excel sheet
Select import columns
Customize your annotation set
Set display options
Your set of annotations will be made available in BiblioSpheres pathway view whenever there are gene ids matching your set in your current BiblioSpheres gene list. You can enter a name and an icon of up to 16 x 16 pixels (8 * 8 pixel images will look best) for your annotation set, or use the default values defined here.
Choose a name for your annotation set:
smoking induc
Upload image file or drag it to the preview area:
Icon preview
Find file
Muster
HSB
RGB
Aktuell:
236, 233, 216
Vorschau
Beispieltext Beispieltext
Beispieltext Beispieltext
Beispieltext Beispieltext
Cancel
< Back
Continue >
Finish

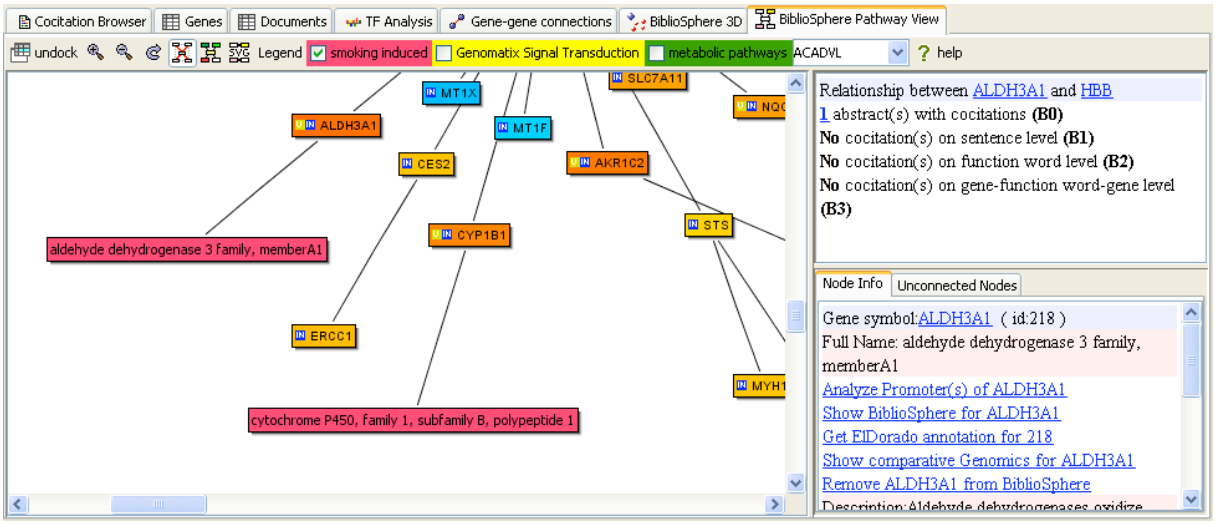
If the selected file is a text file, the Data Import Assistant will look like this:


Data Import Assistant
Steps
Select data import format
Choose Data Separator
Select import columns
Customize your annotation set
The Data Import Assistant tries to recognize the data format of your imported file. Please make sure that the file format was recognized correctly, before you continue.
Original Format
Select the description that fits best for the type of file you want to import:
Delimited -Characters such as commas or tabs separate fields
Fixed Width -Fields have a fixed length, filled with blanks.
Begin import at line
1
Preview of import file:
import data
1 6303 | spermidine/spermine N1-acetyltransferase
2 3081 | "homogentisate 1,2-dioxygenase (homogentisate oxidase)"
3 3417 | "isocitrate dehydrogenase 1 (NADP+), soluble"
4 2512 | "ferritin, light polypeptide"
5 4585 | "mucin 4, tracheobronchial"
6 57402 | S100 calcium binding protein A14
7 6286 | S100 calcium binding protein P
8 2495 | "ferritin, heavy polypeptide 1"
9 6888 | transaldolase 1
10 27075 | transmembrane 4 superfamily member 13
11 10653 | "serine protease inhibitor, Kunitz type, 2"
12 9314 | Kruppel-like factor 4 (gut)
13 1728 | "NAD(P)H dehydrogenase, quinone 1"
14 10551 | anterior gradient 2 homolog (Xenopus laevis)
15 8644 | "aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II)"
16 6748 | "signal sequence receptor, delta (translocon-associated protein delta)"
17 1048 | carcinoembryonic antigen-related cell adhesion molecule 5
Cancel
< Back
Continue >
Finish

In the first step the data format of the file can be selected. The Data Import Assistant will guess the format based on a data sample, and pre-select the values. In the subsequent step, you can set data separators and text qualifiers. Again the Data Import Assistant will make the pre-selection based on the analysis of a sample:


Data Import Assistant
Steps
Select data import format
Choose Data Separator
Select import columns
Customize your annotation set
Choose Data Separator
This Dialog allows you to specify the delimiter used in your imported file. The preview shows you how your file will be processed.
Delimiter
Tab
Semicolon
Comma
Space
Other :
Text Qualifier:
Use first line as column headers
Preview of import file:
0
1
1 6303 spermidine/spermine N1-acetyltransferase
2 3081 homogentisate 1,2-dioxygenase (homogentisa...
3 3417 isocitrate dehydrogenase 1 (NADP+), soluble
4 2512 ferritin, light polypeptide
5 4585 mucin 4, tracheobronchial
6 57402 S100 calcium binding protein A14
7 6286 S100 calcium binding protein P
8 2495 ferritin, heavy polypeptide 1
9 6888 transaldolase 1
10 27075 transmembrane 4 superfamily member 13
11 10653 serine protease inhibitor, Kunitz type, 2
12 9314 Kruppel-like factor 4 (gut)
13 1728 NAD(P)H dehydrogenase, quinone 1
14 10551 anterior gradient 2 homolog (Xenopus laevis)
15 8644 aldo-keto reductase family 1, member C3 (3-al...
16 6748 signal sequence receptor, delta (translocon-as...
17 1048 carcinoembryonic antigen-related cell adhesion...
Cancel
< Back
Continue >
Finish

The following steps are identical with the import of an Excel file. The annotations are made available in the Pathway view of BiblioSphere, just like the metabolic and signal transduction pathways. Whenever one of the genes in BiblioSphere has an annotation, a checkbox will be displayed in the menu bar of this view. All genes that have an annotation in your set carry the icon you selected.



Network Customization

Gene focus

Selecting a gene symbol from the gene selection list in the toolbar redraws the network graph, centred on the selected gene.

Position of gene nodes

You can drag a node to manually change its position. Clicking the Layout Optimization button will reset the layout based on the graph optimization algorithm.

Gene node appearance

You might want to customize the appearance of a node in the network graph, e.g. to highlight a gene of interest. Right-clicking a node will open a context menu. To open a customization dialog, choose the “customize” option. All changes you make will apply to the selected node only.

You can set the following attributes:

Background	Clicking the color bar opens a dialog for selecting the background color
Border	Un-checking this option removes the border from the node box; default is on
Filled	This checkbox toggles the background on/off; default is on
Font	You can change the font, font size, and markup of the displayed text
Icon	Allows you to change the icon displayed in the node box
Shadow	Toggles the node box’s shadow on/off; default is on
Text	Changes the text displayed in the node

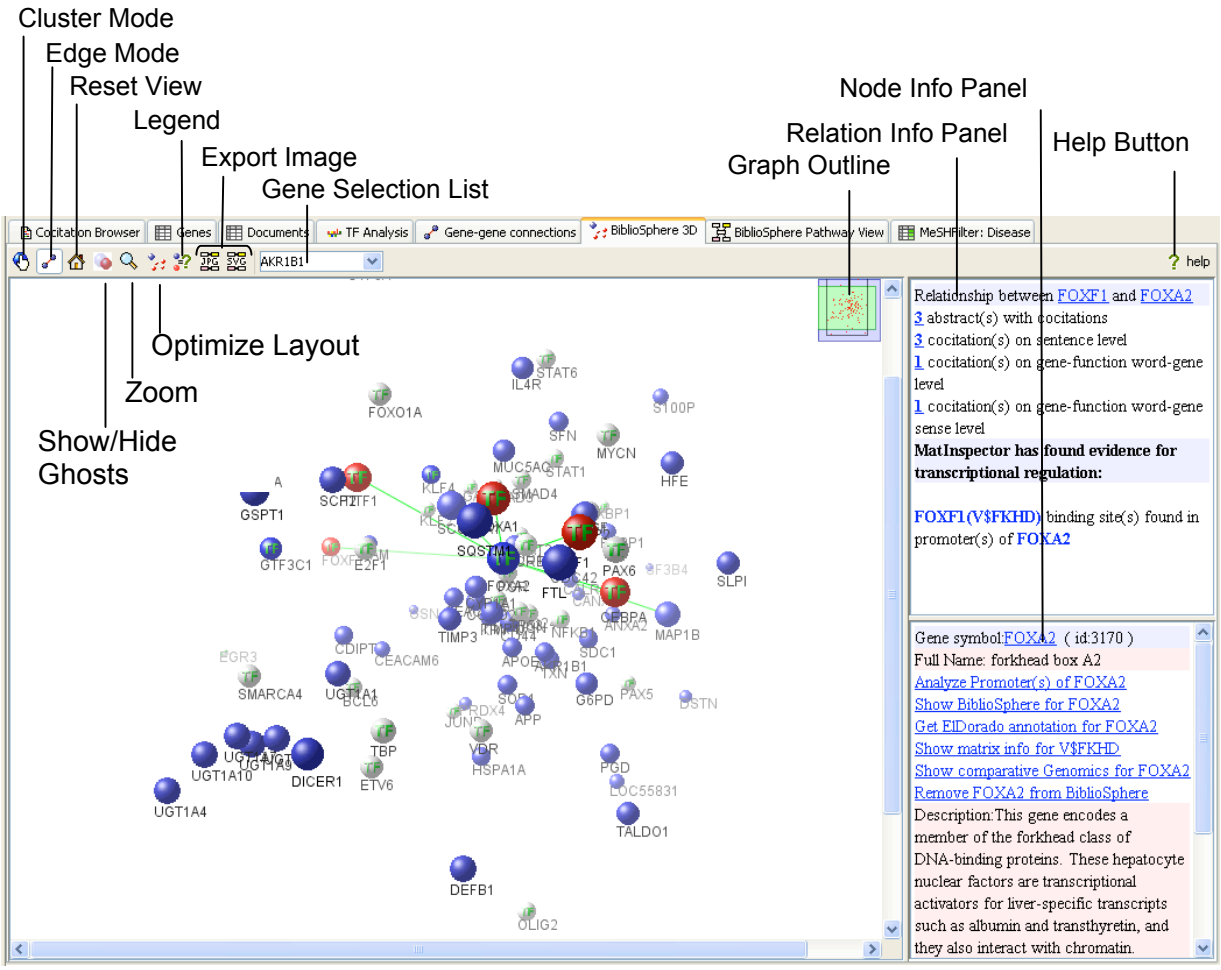
3D View

Overview

The BiblioSphere 3D view lets you navigate the literature in a unique, but intuitive way. Animated 3D graphs allow the identification of complex gene relation structures at first sight.
The 3D View reflects the co-citation frequency as the distance between two genes. The more often two genes occur in the same abstract or sentence, the closer they are in the "Literature Molecule". You can get the information on how often two genes are co-cited by simply moving your mouse pointer over the connecting lines between the spheres that represent these genes. For the lines connecting a gene to become visible, you move the pointer over the gene of interest while the “Show gene connections” option is active. Placing the pointer over a gene displays its full name and Locus ID.

To turn the graph, you can click and drag anywhere in the view pane. Alternatively, you can select a gene from the Gene Selection List to highlight its cluster and bring it to the fore; this is especially handy if you want to get a better look at the environment of a gene in a BiblioSphere containing many elements. For zooming, you click the zoom button, which will drop down a slider that you can move up and down to zoom in and out.

In the 3D view's upper right hand corner you see a miniaturized outline of your BiblioSphere graph. The green area represents the section of the 3D graph that is currently visible to you on the screen, while the black rectangle encloses all spheres in the graph, which are represented as pixel sized red dots. When you open a BiblioSphere, the rectangle is well within the borders of the green area. Zooming into the graph, however, will enlarge the size of the graph; the rectangle then gives you an idea which part of the graph you are currently looking at, and of how to bring spheres that are currently outside of the visible area into focus.


Cluster Mode
Edge Mode
Reset View
Legend
Export Image
Gene Selection List
Node Info Panel
Relation Info Panel
Graph Outline
Help Button
Optimize Layout
Zoom
Show/Hide Ghosts
Cocitation Browser
Genes
Documents
TF Analysis
Gene-gene connections
BiblioSphere 3D
BiblioSphere Pathway View
MeSHFilter: Disease
AKR1B1
help
Relationship between FOXF1 and FOXA2
3 abstract(s) with cocitations
3 cocitation(s) on sentence level
1 cocitation(s) on gene-function word-gene level
1 cocitation(s) on gene-function word-gene sense level
MatInspector has found evidence for transcriptional regulation:
FOXF1(V$FKHD) binding site(s) found in promoter(s) of FOXA2
Gene symbol:FOXA2 (id:3170)
Full Name: forkhead box A2
Analyze Promoter(s) of FOXA2
Show BiblioSphere for FOXA2
Get ElDorado annotation for FOXA2
Show matrix info for V$FKHD
Show comparative Genomics for FOXA2
Remove FOXA2 from BiblioSphere
Description:This gene encodes a member of the forkhead class of DNA-binding proteins. These hepatocyte nuclear factors are transcriptional activators for liver-specific transcripts such as albumin and transthyretin, and they also interact with chromatin.

Toolbar controls and actions:

	Cluster Mode	Allows you to switch to a sub cluster when active, by simply clicking on your cluster centre of choice. Click anywhere in the space between the spheres and connections to switch back to the total survey.
	Edge Mode	Displays edges for a gene when you position the mouse pointer over its sphere.
	Reset View	Resets all filters of your BiblioSphere.
	Show/Hide Ghosts	Ghosts are input genes that do not pass the filter with the current settings
	Zoom	Zoom in and out the 3D view. Clicking the button will drop down a slider that lets you zoom continuously
	Optimize Layout	Recalculate the 3D layout.
	Legend	Display a legend explaining what the different spheres represent.
	Save JPEG	Save a bitmap image of the network in JPEG format.
	Save SVG	Save an image of the network as Scalable Vector Graphic (SVG)

The different types of colored spheres represent functions of genes and the relations between them:

	input gene
	input transcription factor gene
	input gene, filtered out
	input transcription factor gene, filtered out
	co-cited gene
	co-cited transcription factor gene
	gene is co-cited with focused gene
	transcription factor gene is co-cited with focused gene

Information about Genes and Connections

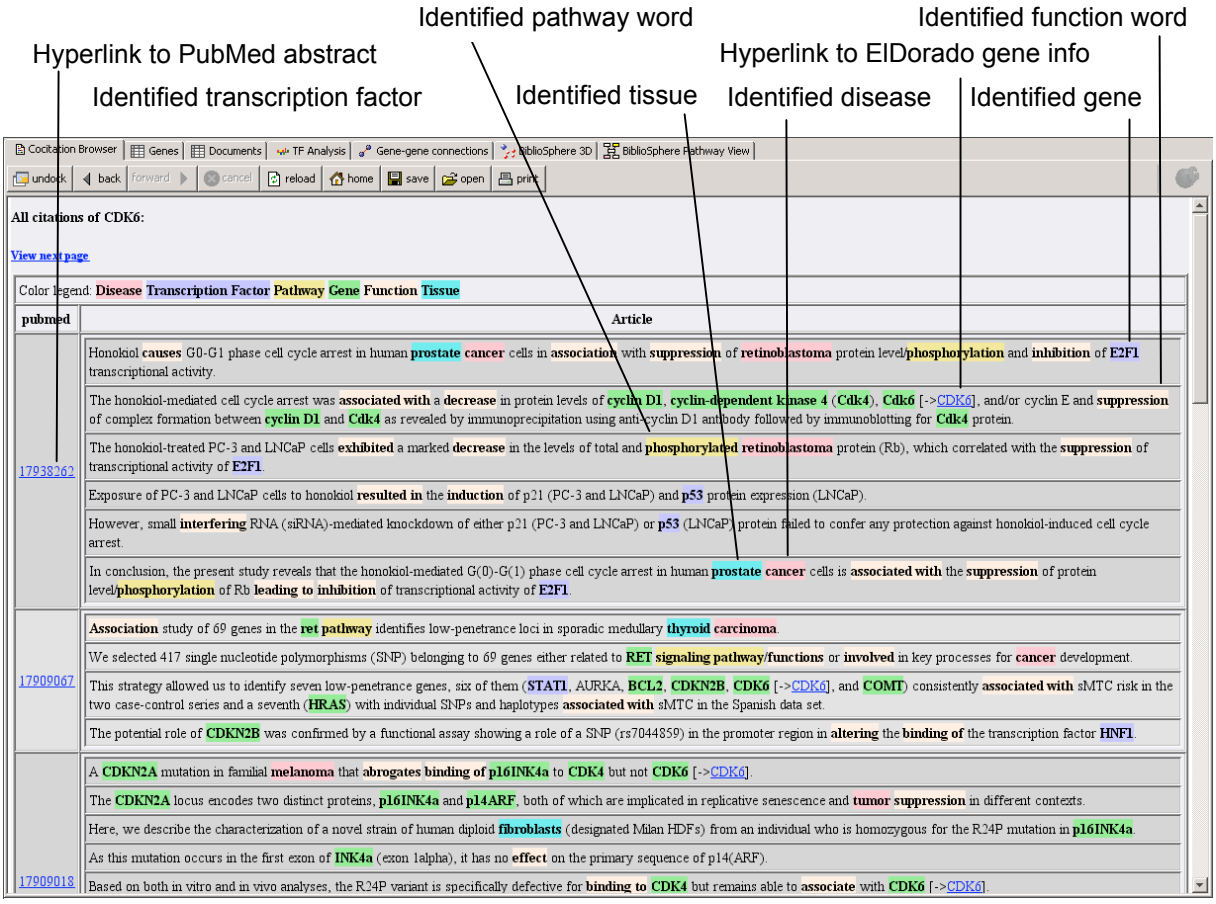
Additional information on focused genes and their relationships is displayed in the Info Panels next to the 3D view (see Relation Info Panel and Node Info Panel for details; in contrast to the BiblioSphere Pathway View, the Node Info Panel here is not subdivided into tabs, as there is no list of unconnected nodes). If one of these genes is a transcription factor, you can additionally gain information on potential binding sites in the promoters of its cocited neighbours. Connections affirmed by this promoter analysis are displayed in green.

Focus on Gene Subnets (Clusters)

Activating the Expand/Collapse Gene Cluster option adds yet another level of dynamic filtering to your BiblioSphere: By clicking on a sphere you hide all elements that are not directly connected to it from view, leaving only the cluster around the selected gene visible. The graph's zoom factor is adjusted automatically to provide an optimized view on the region of interest. Clicking anywhere between the displayed elements lets you see the whole graph again. Furthermore, you can activate the Show Ghosts option, which will show faded those input gene spheres that are blocked from view by the current combination of filters.

Co-Citation Browser

The Co-citation Browser provides access to the citations of a gene in the literature.
You display citations of a gene in the browser by clicking on the row header of a gene dataset in the Genes View table; all citations of that gene will be shown. Alternatively, you can view co-citations of two genes. To that aim, click on the appropriate co-citation link in the Relation Info Panel of the Pathway View or the 3D View.


Identified pathway word
Identified function word
Hyperlink to PubMed abstract
Hyperlink to ElDorado gene info
Identified transcription factor
Identified tissue
Identified disease
Identified gene
Cocitation Browser
Genes
Documents
TF Analysis
Gene-gene connections
BiblioSphere 3D
BiblioSphere Pathway View
undock
back
forward
cancel
reload
home
save
open
print
All citations of CDK6:
View next page
Color legend: Disease Transcription Factor Pathway Gene Function Tissue
pubmed
Article
17938262
Honokiol causes G0-G1 phase cell cycle arrest in human prostate cancer cells in association with suppression of retinoblastoma protein level/phosphorylation and inhibition of E2F1 transcriptional activity.
The honokiol-mediated cell cycle arrest was associated with a decrease in protein levels of cyclin D1, cyclin-dependent kinase 4 (Cdk4), Cdk6 [->CDK6], and/or cyclin E and suppression of complex formation between cyclin D1 and Cdk4 as revealed by immunoprecipitation using anti-cyclin D1 antibody followed by immunoblotting for Cdk4 protein.
The honokiol-treated PC-3 and LNCaP cells exhibited a marked decrease in the levels of total and phosphorylated retinoblastoma protein (Rb), which correlated with the suppression of transcriptional activity of E2F1.
Exposure of PC-3 and LNCaP cells to honokiol resulted in the induction of p21 (PC-3 and LNCaP) and p53 protein expression (LNCaP).
However, small interfering RNA (siRNA)-mediated knockdown of either p21 (PC-3 and LNCaP) or p53 (LNCaP) protein failed to confer any protection against honokiol-induced cell cycle arrest.
In conclusion, the present study reveals that the honokiol-mediated G(0)-G(1) phase cell cycle arrest in human prostate cancer cells is associated with the suppression of protein level/phosphorylation of Rb leading to inhibition of transcriptional activity of E2F1.
17909067
Association study of 69 genes in the ret pathway identifies low-penetrance loci in sporadic medullary thyroid carcinoma.
We selected 417 single nucleotide polymorphisms (SNP) belonging to 69 genes either related to RET signaling pathway/functions or involved in key processes for cancer development.
This strategy allowed us to identify seven low-penetrance genes, six of them (STAT1, AURKA, BCL2, CDKN2B, CDK6 [->CDK6], and COMT) consistently associated with sMTC risk in the two case-control series and a seventh (HRAS) with individual SNPs and haplotypes associated with sMTC in the Spanish data set.
The potential role of CDKN2B was confirmed by a functional assay showing a role of a SNP (rs7044859) in the promoter region in altering the binding of the transcription factor HNF1.
17909018
A CDKN2A mutation in familial melanoma that abrogates binding of p16INK4a to CDK4 but not CDK6 [->CDK6].
The CDKN2A locus encodes two distinct proteins, p16INK4a and p14ARF, both of which are implicated in replicative senescence and tumor suppression in different contexts.
Here, we describe the characterization of a novel strain of human diploid fibroblasts (designated Milan HDFs) from an individual who is homozygous for the R24P mutation in p16INK4a.
As this mutation occurs in the first exon of INK4a (exon 1alpha), it has no effect on the primary sequence of p14(ARF).
Based on both in vitro and in vivo analyses, the R24P variant is specifically defective for binding to CDK4 but remains able to associate with CDK6 [->CDK6].

Link to PubMed

The browser displays the PubMed IDs of the relevant articles as links to the PubMed abstracts, which will be displayed in an external browser if you click on the link.

Tagged Sentences

Every sentence in the abstract that cites the gene, or in the case of co-citation abstracts, one of the co-cited genes, is displayed for a quick overview. Expressions identified as denoting a transcription factor, a gene, a tissue, a disease, a function word, or a pathway associated term, are color tagged to facilitate assessment of the context. For co-citations on abstract level, every sentence in the abstract citing any gene is displayed.

Table Views

Overview

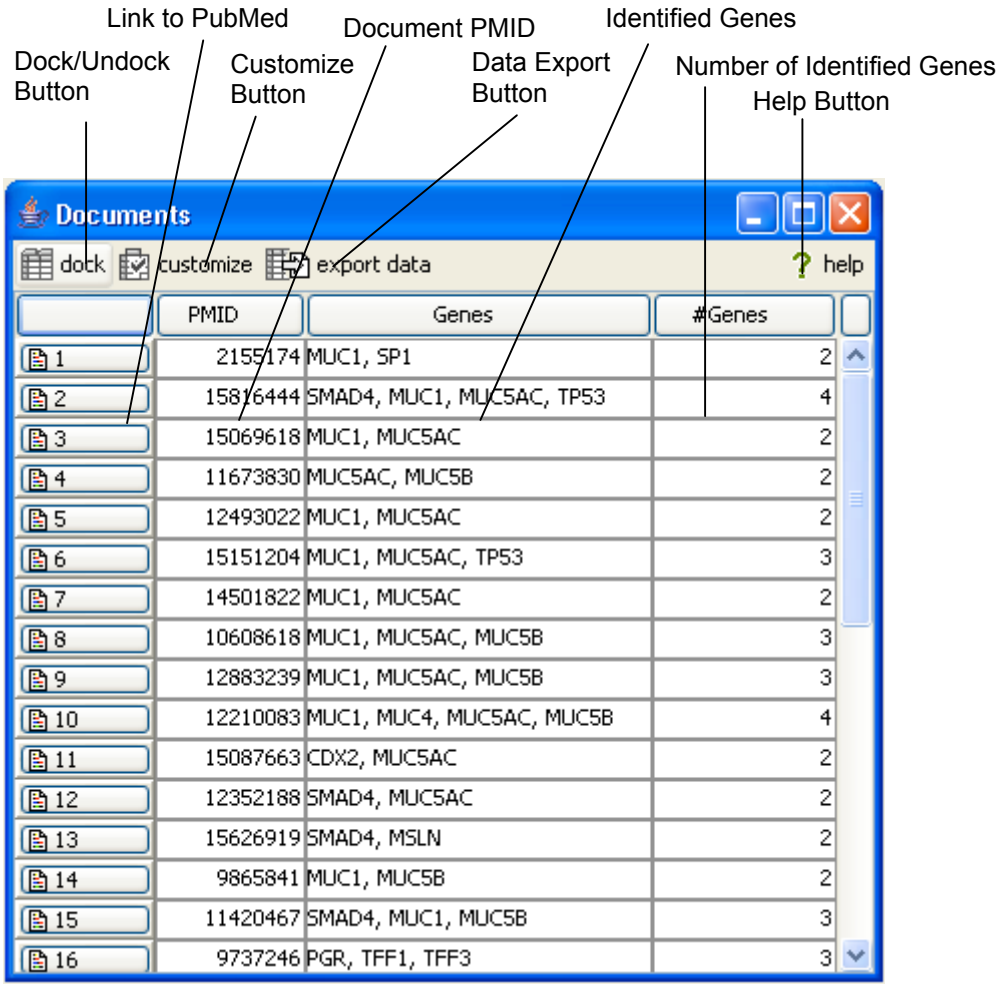
BSPE offers tabular views to the data in your BiblioSphere, specifically to genes, gene-gene connections, documents containing references to the genes, and transcription factor analysis.
You can sort any table by any column, select/deselect display of individual columns, and export the data to an Excel file.

Toolbar controls and actions:

Dock/Undock	Open table in a separate window/return table to BiblioSphere window
Customize	Hide/show individual columns in the table
Export Data	Save the table in Excel format
Help	Display the BiblioSphere help

Documents Table

The Documents view component displays information for the PubMed abstracts compiled into your BiblioSphere which pass the filter at the current settings, and links directly to PubMed. For each document, the PMID, the identified genes, and their number are displayed.
Clicking the row number opens the relevant PubMed article in an external browser.



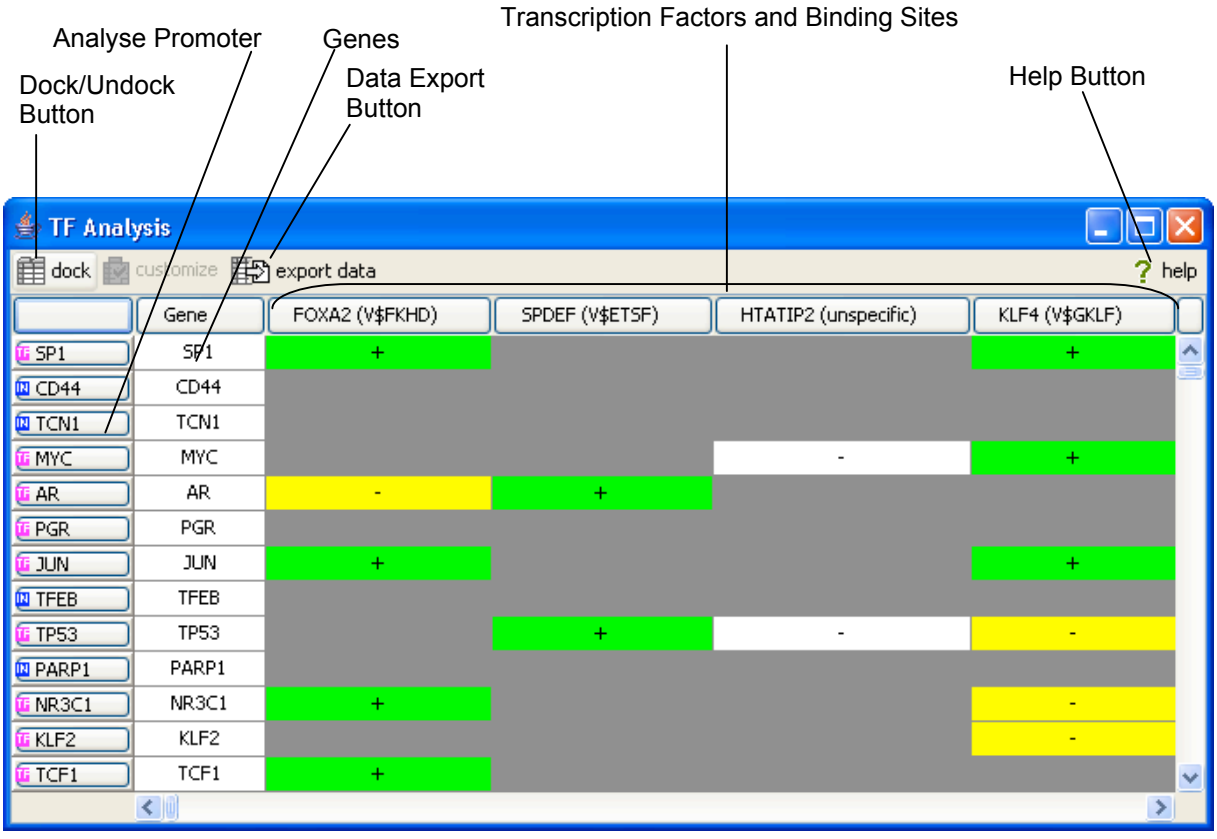
	PMID	Genes	#Genes
1	2155174	MUC1, SP1	2
2	15816444	SMAD4, MUC1, MUC5AC, TP53	4
3	15069618	MUC1, MUC5AC	2
4	11673830	MUC5AC, MUC5B	2
5	12493022	MUC1, MUC5AC	2
6	15151204	MUC1, MUC5AC, TP53	3
7	14501822	MUC1, MUC5AC	2
8	10608618	MUC1, MUC5AC, MUC5B	3
9	12883239	MUC1, MUC5AC, MUC5B	3
10	12210083	MUC1, MUC4, MUC5AC, MUC5B	4
11	15087663	CDX2, MUC5AC	2
12	12352188	SMAD4, MUC5AC	2
13	15626919	SMAD4, MSLN	2
14	9865841	MUC1, MUC5B	2
15	11420467	SMAD4, MUC1, MUC5B	3
16	9737246	PGR, TFF1, TFF3	3

TF Analysis

The TF Analysis component displays the results of MatInspector™ analysis for transcription factor binding sites in the promoters of co-cited genes in your current selection of BiblioSphere data, and provides links to promoter analysis with GEMS Launcher.
Promoters are checked for binding sites of transcription factors present in the BiblioSphere if a MatInspector™ matrix is available. Each row in the table represents one gene promoter analysis result. Each column represents the results for one transcription factor. The meanings of the cell entries are as follows:

+	The co-cited transcription factor has a binding site in at least one of this gene's alternative promoters.
-	No binding site for this co-cited transcription factor was found.
-	The gene is co-cited, but a matrix is not yet available for this transcription factor.
	The analyzed gene and transcription factor were not found co-cited.

Clicking the "Analyse Promoter" button of a gene displays the analysis of its promoter with GEMS Launcher in an external browser. Input genes are marked with the IN symbol, transcription factors with the TF icon.



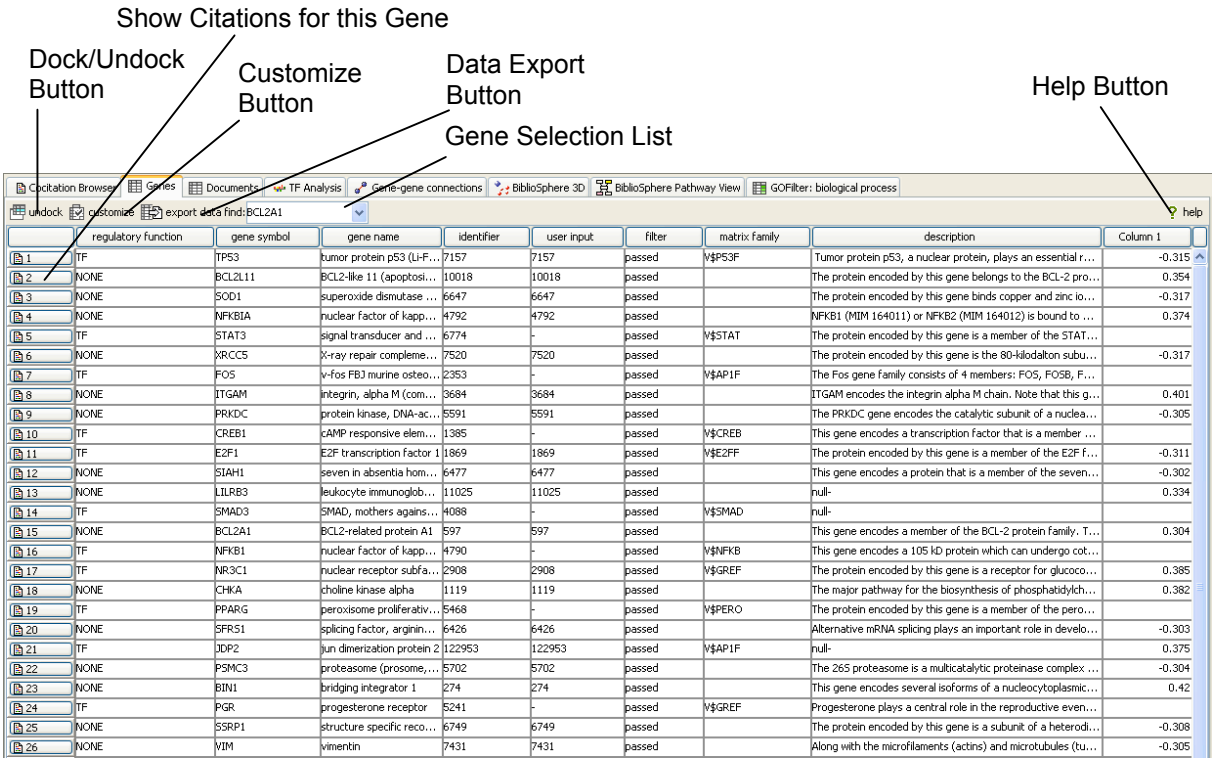
Genes

The Genes spreadsheet displays all the genes in your current selection of BiblioSphere data.
You can jump directly to a gene of interest in the list by selecting a gene symbol in the Gene Selection List in the toolbar.

A click on the row header of a gene dataset opens the Cocitation Browser displaying the citations for this gene.

Table Content:

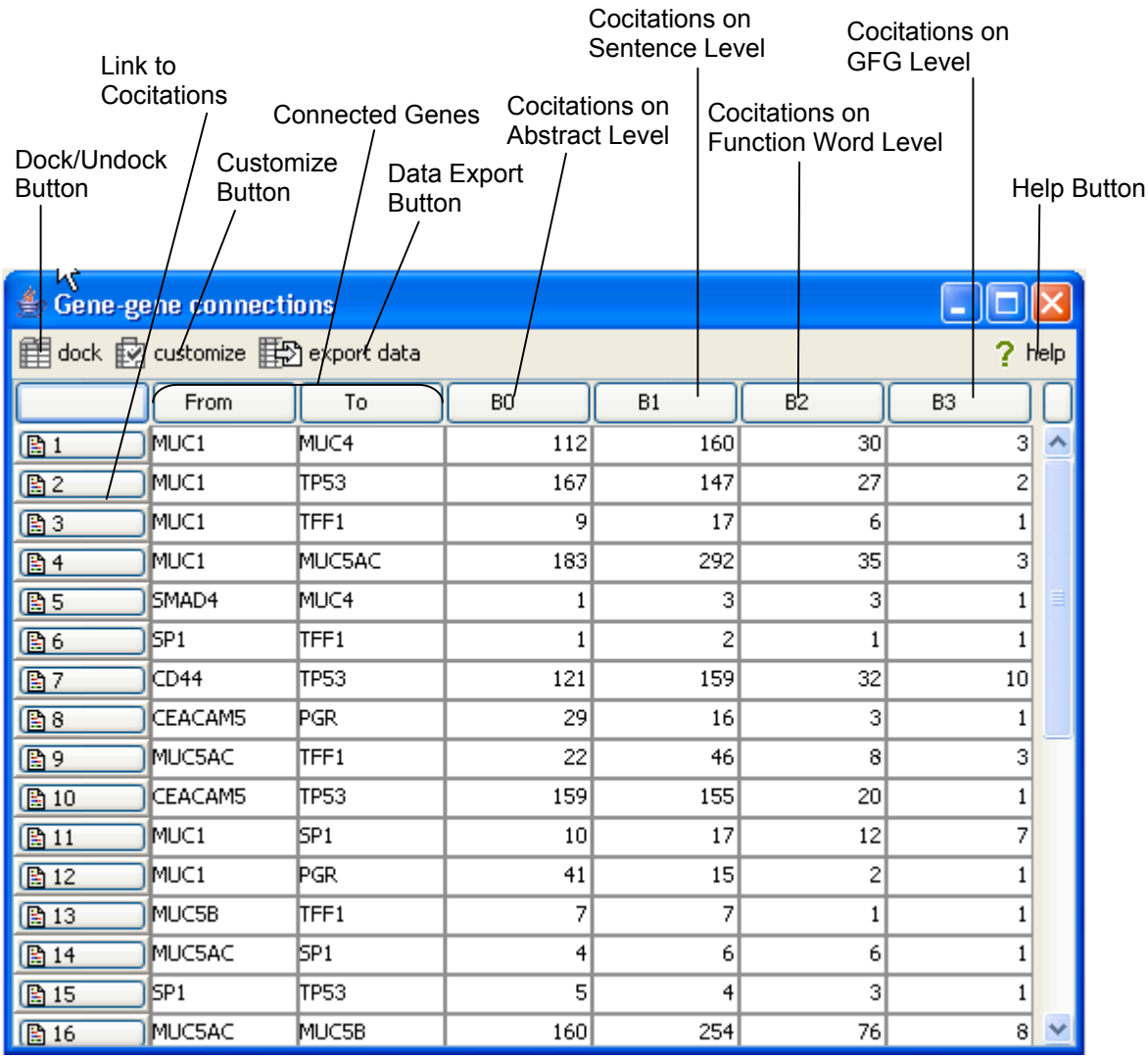
Column Name	Content
Row Header	The row header as a specialized column links directly to co-citations of the gene in this row.
Regulatory Function	Shows Genomatix annotation for genes with a known regulatory function. Currently only transcription factors are annotated.
Gene Symbol	The official or preferred symbol of this gene.
Gene Name	The official or preferred name of this gene.
Identifier	The gene identifier (Gene ID).
User Input	Shows the original query term entered by the user to find this gene.
Filter	Indicates whether the gene has passed the filter or has been blocked. Blocked genes are not displayed in the network views
Matrix Family	For transcription factors, matrix families from the MatInspector library are displayed here.
Description	A description of the gene, as provided by NCBI.
User Data	Expression value for each gene, if provided for analysis by the user.



	regulatory function	gene symbol	gene name	identifier	user input	filter	matrix family	description	Column 1
1	TF	TP53	tumor protein p53 (Li-F…	7157	7157	passed	V$P53F	Tumor protein p53, a nuclear protein, plays an essential r…	-0.315
2	NONE	BCL2L11	BCL2-like 11 (apoptosi…	10018	10018	passed		The protein encoded by this gene belongs to the BCL-2 pro…	0.354
3	NONE	SOD1	superoxide dismutase …	6647	6647	passed		The protein encoded by this gene binds copper and zinc io…	-0.317
4	NONE	NFKBIA	nuclear factor of kapp…	4792	4792	passed		NFKB1 (MIM 164011) or NFKB2 (MIM 164012) is bound to …	0.374
5	TF	STAT3	signal transducer and …	6774	-	passed	V$STAT	The protein encoded by this gene is a member of the STAT…	
6	NONE	XRCC5	X-ray repair compleme…	7520	7520	passed		The protein encoded by this gene is the 80-kilodalton subu…	-0.317
7	TF	FOS	v-fos FBJ murine osteo…	2353	-	passed	V$AP1F	The Fos gene family consists of 4 members: FOS, FOSB, F…	
8	NONE	ITGAM	integrin, alpha M (com…	3684	3684	passed		ITGAM encodes the integrin alpha M chain. Note that this g…	0.401
9	NONE	PRKDC	protein kinase, DNA-ac…	5591	5591	passed		The PRKDC gene encodes the catalytic subunit of a nuclea…	-0.305
10	TF	CREB1	cAMP responsive elem…	1385	-	passed	V$CREB	This gene encodes a transcription factor that is a member …	
11	TF	E2F1	E2F transcription factor 1	1869	1869	passed	V$E2FF	The protein encoded by this gene is a member of the E2F f…	-0.311
12	NONE	SIAH1	seven in absentia hom…	6477	6477	passed		This gene encodes a protein that is a member of the seven…	-0.302
13	NONE	LILRB3	leukocyte immunoglob…	11025	11025	passed		null-	0.334
14	TF	SMAD3	SMAD, mothers agains…	4088	-	passed	V$SMAD	null-	
15	NONE	BCL2A1	BCL2-related protein A1	597	597	passed		This gene encodes a member of the BCL-2 protein family. T…	0.304
16	TF	NFKB1	nuclear factor of kapp…	4790	-	passed	V$NFKB	This gene encodes a 105 kD protein which can undergo cot…	
17	TF	NR3C1	nuclear receptor subfa…	2908	2908	passed	V$GREF	The protein encoded by this gene is a receptor for glucoco…	0.385
18	NONE	CHKA	choline kinase alpha	1119	1119	passed		The major pathway for the biosynthesis of phosphatidylch…	0.382
19	TF	PPARG	peroxisome proliferativ…	5468	-	passed	V$PERO	The protein encoded by this gene is a member of the pero…	
20	NONE	SFRS1	splicing factor, arginin…	6426	6426	passed		Alternative mRNA splicing plays an important role in develo…	-0.303
21	TF	JDP2	jun dimerization protein 2	122953	122953	passed	V$AP1F	null-	0.375
22	NONE	PSMC3	proteasome (prosome,…	5702	5702	passed		The 26S proteasome is a multicatalytic proteinase complex …	-0.304
23	NONE	BIN1	bridging integrator 1	274	274	passed		This gene encodes several isoforms of a nucleocytoplasmic…	0.42
24	TF	PGR	progesterone receptor	5241	-	passed	V$GREF	Progesterone plays a central role in the reproductive even…	
25	NONE	SSRP1	structure specific reco…	6749	6749	passed		The protein encoded by this gene is a subunit of a heterodi…	-0.308
26	NONE	VIM	vimentin	7431	7431	passed		Along with the microfilaments (actins) and microtubules (tu…	-0.305

Gene-Gene Connections

The spreadsheet view of co-citations contained in a BiblioSphere contains direct links to the Cocitation Browser and supplies the functionality to export data for further analysis with external applications. While the BiblioSphere 3D view displays the edges of one gene at a time, the Gen-Gene Connections spreadsheet holds all literature based relations in your current selection of BiblioSphere data.

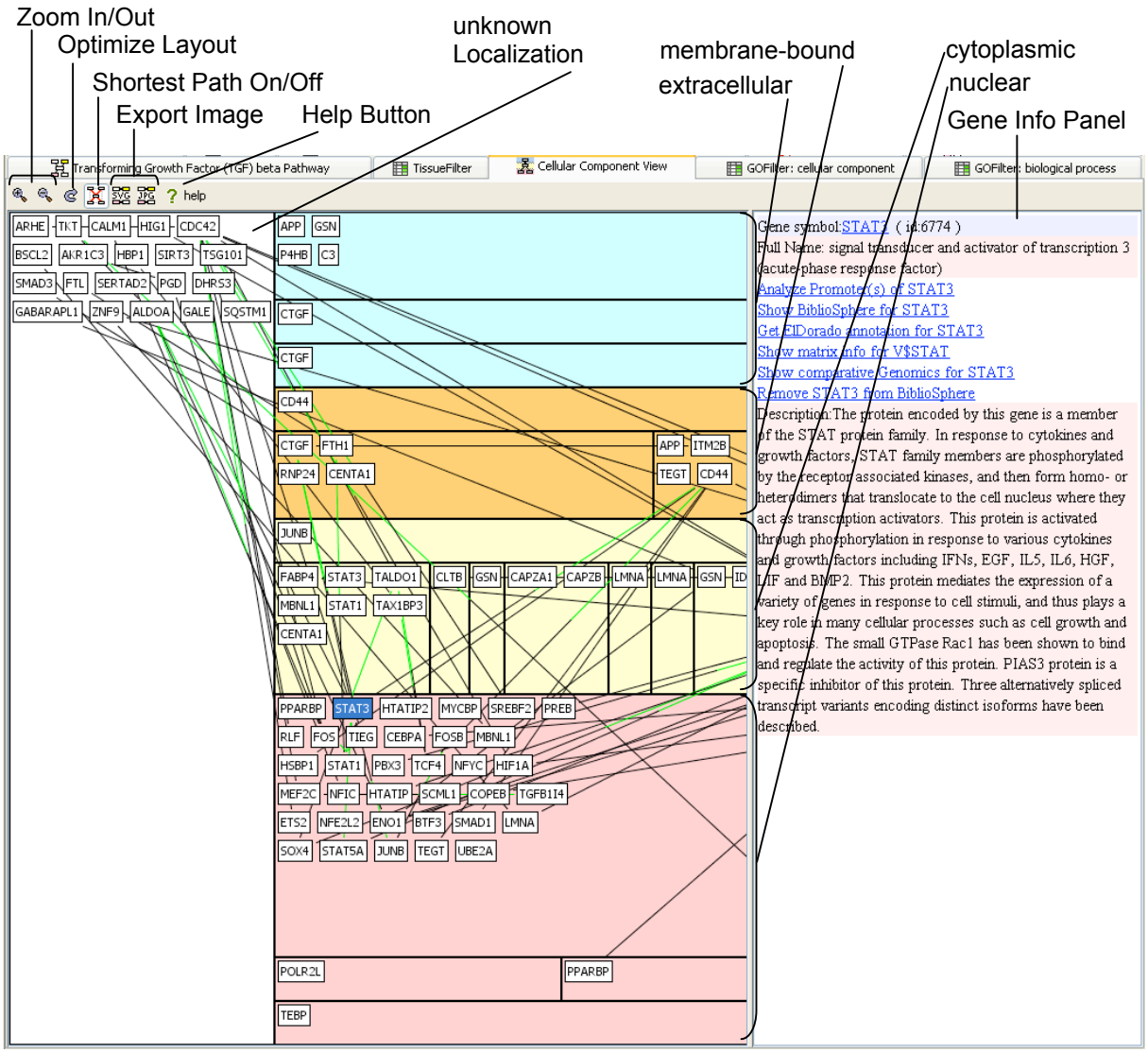


	From	To	B0	B1	B2	B3
1	MUC1	MUC4	112	160	30	3
2	MUC1	TP53	167	147	27	2
3	MUC1	TFF1	9	17	6	1
4	MUC1	MUC5AC	183	292	35	3
5	SMAD4	MUC4	1	3	3	1
6	SP1	TFF1	1	2	1	1
7	CD44	TP53	121	159	32	10
8	CEACAM5	PGR	29	16	3	1
9	MUC5AC	TFF1	22	46	8	3
10	CEACAM5	TP53	159	155	20	1
11	MUC1	SP1	10	17	12	7
12	MUC1	PGR	41	15	2	1
13	MUC5B	TFF1	7	7	1	1
14	MUC5AC	SP1	4	6	6	1
15	SP1	TP53	5	4	3	1
16	MUC5AC	MUC5B	160	254	76	8

Cellular Component View

The Cellular Component View helps you to identify the subcellular compartments relevant to your set of genes.
The hierarchical Gene Ontology filter assigns genes to their subcellular compartment. Based on these annotations, the Cellular Component View builds a diagrammed cell layout. Therefore the Cellular Component View is only available if the GO filter "Cellular Component" has been activated. Network genes are displayed in their subcellular location helping to reconstruct the way a signal takes through the cell compartments during a regulatory event.
The basic localizations are color-coded, with blue denoting extra-cellular, orange, membrane-bound, yellow, cytoplasmic, pink, nuclear, and white, unknown localization. Gene symbols are subgrouped further according to sub-compartmentalization.

Basic Navigation
Nodes can be dragged with the mouse pointer to customize the layout. A single click on a gene node displays summary info in the Node Info Panel, which shows detailed information on the selected gene and provides links for further analysis. The shortest path for each gene can be brought up by a double click on it in the Cellular Component View Panel.

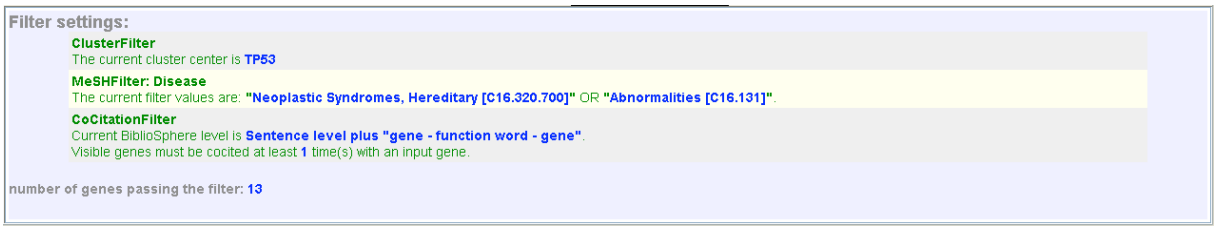

Zoom In/Out
Optimize Layout
Shortest Path On/Off
Export Image
Help Button
unknown
Localization
membrane-bound
extracellular
cytoplasmic
nuclear
Gene Info Panel
Transforming Growth Factor (TGF) beta Pathway
TissueFilter
Cellular Component View
GOFilter: cellular component
GOFilter: biological process
help
Gene symbol:STAT1 (id:6774)
Full Name: signal transducer and activator of transcription 3 (acute-phase response factor)
Analyze Promoter(s) of STAT3
Show BiblioSphere for STAT3
Get ElDorado annotation for STAT3
Show matrix info for V$STAT
Show comparative Genomics for STAT3
Remove STAT3 from BiblioSphere
Description:The protein encoded by this gene is a member of the STAT protein family. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor-associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators. This protein is activated through phosphorylation in response to various cytokines and growth factors including IFNs, EGF, IL5, IL6, HGF, LIF and BMP2. This protein mediates the expression of a variety of genes in response to cell stimuli, and thus plays a key role in many cellular processes such as cell growth and apoptosis. The small GTPase Rac1 has been shown to bind and regulate the activity of this protein. PIAS3 protein is a specific inhibitor of this protein. Three alternatively spliced transcript variants encoding distinct isoforms have been described.

Toolbar Controls and Actions

	Zoom In	Zoom in to the view
	Zoom Out	Zoom out
	Optimize Layout	Recalculate the layout depending on the gene in focus
	Shortest Path	Shortest path for the gene in focus ON/OFF
	Save SVG	Save an image of the Network as Scalable Vector Graphic (SVG)
	Save JPEG	Save a bitmap image of the Network in JPEG format.

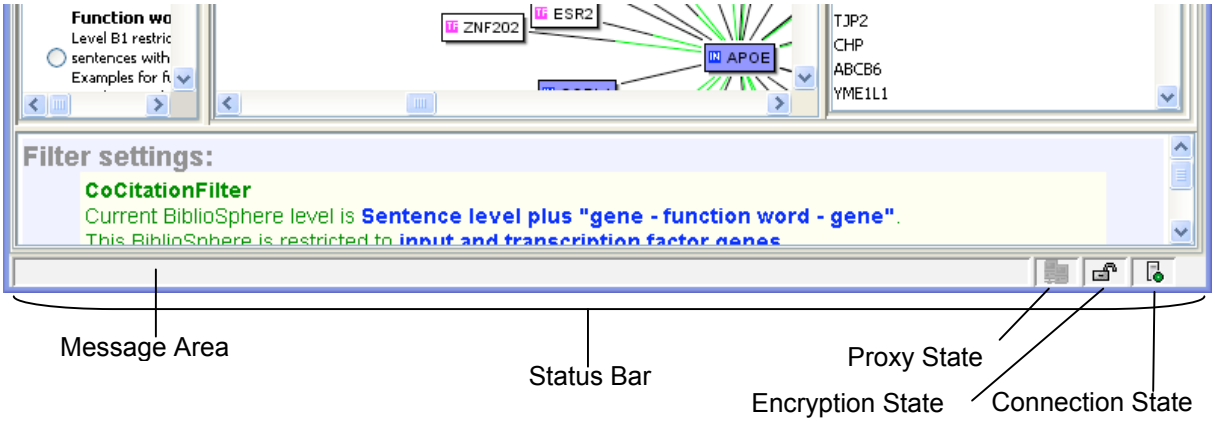
Protocol Panel

The Protocol Panel displays the current filter settings and the number of genes passing the filter.


Filter settings:
ClusterFilter
The current cluster center is TP53
MeSHFilter: Disease
The current filter values are: "Neoplastic Syndromes, Hereditary [C16.320.700]" OR "Abnormalities [C16.131]".
CoCitationFilter
Current BiblioSphere level is Sentence level plus "gene - function word - gene".
Visible genes must be cocited at least 1 time(s) with an input gene.
number of genes passing the filter: 13

Status Bar

The Status Bar keeps you informed about your internet connection settings and state.



Status Bar Items:

Message Area		Information on the state of the application is displayed here.
Proxy State		The colored icon indicates that a proxy server is used to connect to the internet.
		The icon is greyed out when no proxy server is used.
Encryption State		Connection encrypted (SSL).
		No encryption is used.
Connection State		The client is not connected to a BiblioSphere server on the internet
		Trying to connect.
		Connected to a BiblioSphere server.

Network Filtering

Overview

The unfiltered output of any literature mining tool often includes large portions of data that may be only marginally relevant to the user's focus of interest. BSPE offers a potent system of filters to customize the analysis output according to your needs. This includes filters based on the content of the literature itself, as well as on functional analysis using hierarchical annotation terms. A statistical evaluation of the results of the functional analysis is available to facilitate focussing on the most relevant findings.

Filter Panel

The Filter Panel contains all active filters for a BiblioSphere. The Free Text and Co-Citation Filters are always available. Additionally, you can load biological entity filters based on MeSH or Gene Ontology terms, or on UniGene tissue names. To this end, you select the desired filter from the Filter menu. Loaded filters are check marked in the menu. To unload and thus deactivate a filter, uncheck the according menu item. Filters act additively; you can load and activate any number of them in parallel.

Changes in the filter settings will affect the content of the following views:

- Gene List
- Documents List
- TF Analysis
- Gene-gene Connection List
- 3D View
- Pathway View

Literature Analysis Filter

Co-Citation Filter

The Co-Citation Filter filters your BiblioSphere data based on co-citation frequency and semantic specificity levels.
BiblioSpheres from free text, MeSH term, or PMID list search analyses contain all genes found in the search. However, if your BiblioSphere was generated in an analysis based on gene identifiers that you provided by manual input or file upload, you can specify what kinds of genes you want to see:

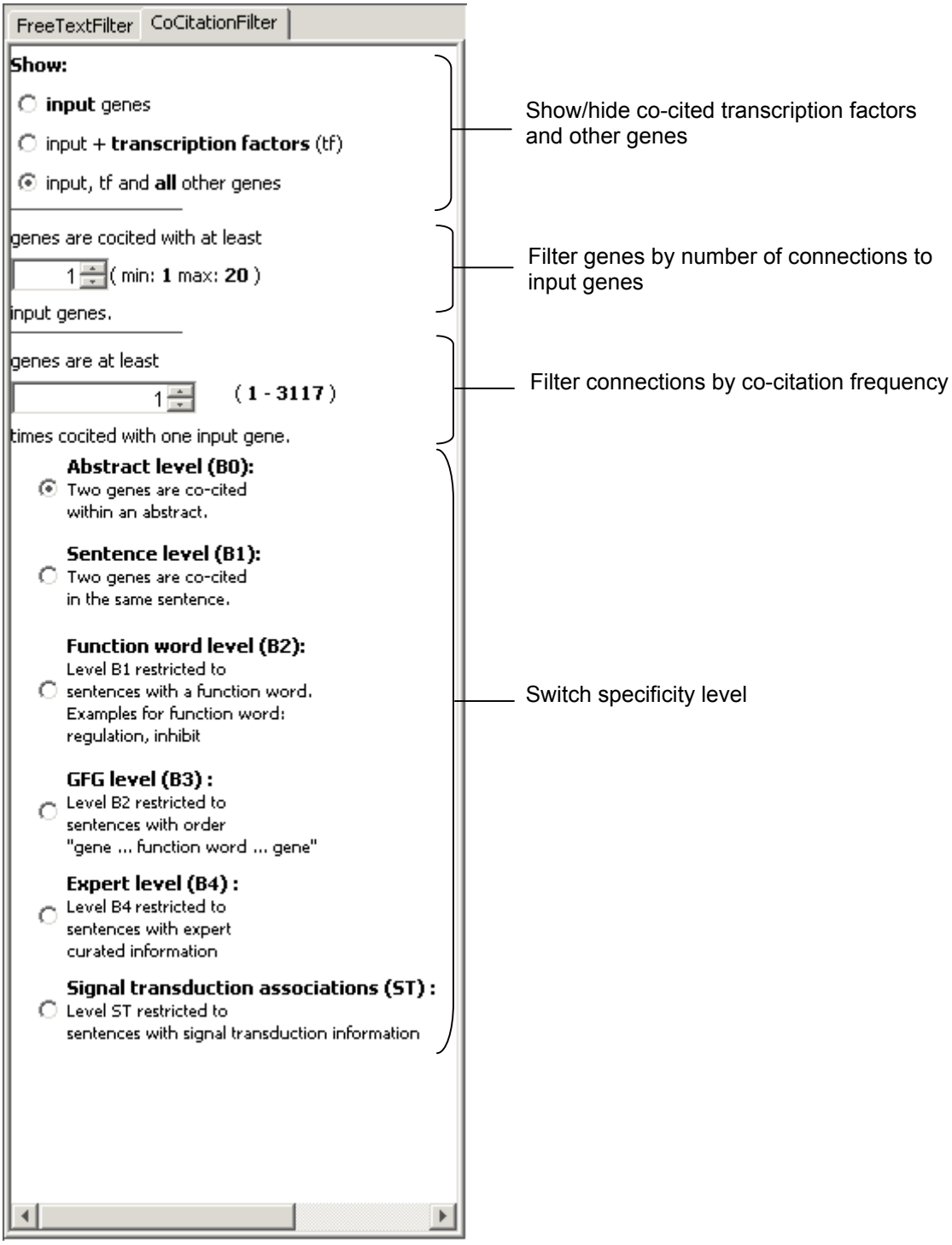
- Only your input genes
- Your input genes and the transcription factor genes connected to them
- Your input genes and all connected genes

The last option is available if you entered the gene identifiers manually and unchecked the "Show only co-cited transcription factor genes" option in the analysis. An analysis based on uploaded data is restricted to finding co-cited transcription factors.

If you choose the second or third option, further filtering options pertaining to the non-input genes become available:

- In a CCBS, you can select the number of input genes another gene has to be connected to in order to be displayed (in an SGBS, this number will always be 1); the maximum is the total number of input genes.
- You can choose the number of times a gene has to be co-cited with one input gene; the maximum is the largest number of co-citations for a non-input gene with an input gene in your BiblioSphere.

The specificity level options are always available.


FreeTextFilter
CoCitationFilter
Show:
input genes
input + transcription factors (tf)
input, tf and all other genes
genes are cocited with at least
1 (min: 1 max: 20)
input genes.
genes are at least
1 (1 - 3117)
times cocited with one input gene.
Abstract level (B0):
Two genes are co-cited within an abstract.
Sentence level (B1):
Two genes are co-cited in the same sentence.
Function word level (B2):
Level B1 restricted to sentences with a function word. Examples for function word: regulation, inhibit
GFG level (B3) :
Level B2 restricted to sentences with order "gene ... function word ... gene"
Expert level (B4) :
Level B4 restricted to sentences with expert curated information
Signal transduction associations (ST) :
Level ST restricted to sentences with signal transduction information
Show/hide co-cited transcription factors and other genes
Filter genes by number of connections to input genes
Filter connections by co-citation frequency
Switch specificity level

Specificity Levels

Genomatix BiblioSphere includes six different filter levels for gene-gene-co-citations. As an example, co-citations of the transcription factors E2F1 and TP53 are shown for each of the six levels. Transcription factor names are printed in indigo, other gene names in green, tissues in cyan, and function words in pink.

Abstract level:
E2F1 and TP53 are co-cited somewhere within an abstract of a publication:

17321453	The **epithelial** component was strongly positive for **CK13**, **CK14**, and AE1/AE, and groups of positive cells were seen for **CK19** and **vimentin**.
	The whole mesenchymal component was positive for **vimentin**, negative for cytokeratins, and focal cells were positive for smooth- **muscle** actin.
	Both components were strongly positive for **P53** [->TP53] and **Cyclin D1**, and focally positive for **MDM2**.
	Rare multinucleated **giant cells** showed expression of **CD68**, and focal dendritical cells on carcinomatous nests were positive for S-100.
	The **CK7**, **CK8**, **Factor XIIIa**, c-erbB-2, P16, **CDK-4**, **Rb1**, and **E2F-1** [->E2F1] were not detected in these 2 groups of malignant cell populations.

Sentence level:
E2F1 and TP53 are co-cited within the same sentence:

11536046	In addition, K5 Myc and K5 **E2F1** [->E2F1] transgenic mice both display aberrant, **p53** [->TP53]-dependent apoptosis in the **epidermis**.

Sentence level plus "function word":
E2F1 and TP53 are co-cited in the same sentence and the sentence also contains a "function word" (colored in light pink). Examples of function words are: regulation, inhibit, modulate, enhance:

9519869	Deregulation of the S-phase **promoting** **E2F-1** [->E2F1] transcription factor has been shown to cooperate with **p53** [->TP53] to **induce** apoptosis.

Sentence level plus "gene - function word - gene":
E2F1 and TP53 are co-cited in one sentence and connected via a "function word".

12625370	**E2F1** [->E2F1] **activates** the human **p53** [->TP53] promoter and overcomes the **repressive** **effect** of **hepatitis B** viral X protein (Hbx) on the **p53** [->TP53] promoter.
	E2F1 [->E2F1] **activated** the **p53** [->TP53] promoter through **E2F1** [->E2F1] binding site.
	When **E2F1** [->E2F1] was cotransfected with HBx, **E2F1** [->E2F1] overcame the **repressive** **effect** of HBx on the **p53** [->TP53] promoter through the **E2F1** [->E2F1] site.

Expert level:
Hand annotated co-citation of E2F1 and TP53.

15255934	**Overexpression** of **E2F-1** [->E2F1] in PC12 cells was sufficient to **induce** **p53** [->TP53] and **Bax** proteins, as well as nuclear fragmentation.

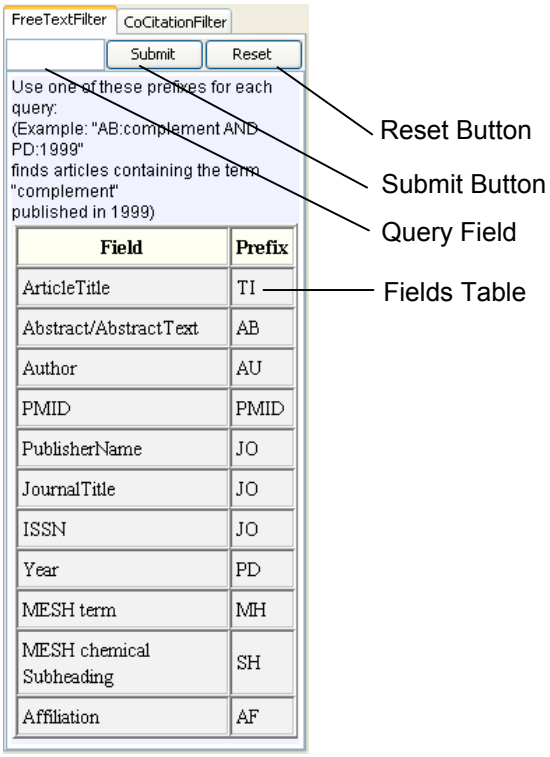
Signal transduction associations:
E2F1 and TP53 are co-cited in a sentence containing a pathway-associated term (ochre background); at least one of the co-cited genes bears a Genomatix signal transduction pathway annotation.

16582589	Myc and **E2F1** [->E2F1] engage the ATM **signaling** **pathway** to **activate** **p53** [->TP53] and **induce** apoptosis.

Implications:
The Abstract level comprises all gene-gene-co-citations available from the literature without ignoring any information. The advantage of this level is the broad statistical basis. The advantage of the other filter levels is the increasing specificity.

Free Text Filter

The Free Text Filter filters the documents that make up your BiblioSphere by a full text search with your query terms.
Use prefixes separated by a colon from your search term to specify the document fields you want to search in. Available document fields and prefixes are listed in the fields table:



Field	Prefix
ArticleTitle	TI
Abstract/AbstractText	AB
Author	AU
PMID	PMID
PublisherName	JO
JournalTitle	JO
ISSN	JO
Year	PD
MESH term	MH
MESH chemical Subheading	SH
Affiliation	AF

Biological Entity Filter

Overview

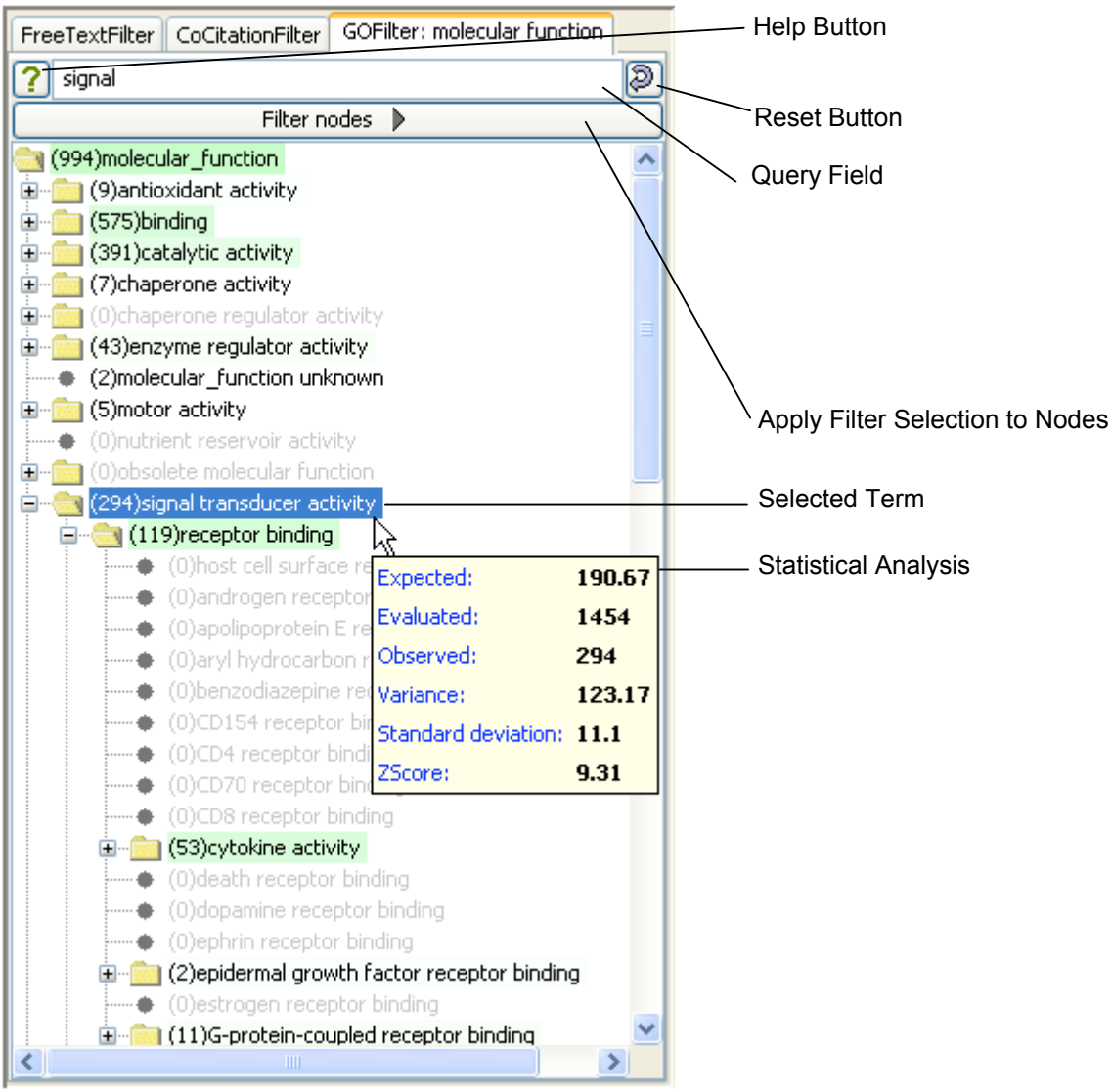
Filtering

In any of the hierarchical filters you can enter a term in the "Query Field" and press return. This will expand the tree to show the first matching term, or you can select your filter term by clicking on it. Pressing the "control" key during selection enables you to use a combination of terms for filtering. These filter terms are combined using the OR operator. The current combination of filters is displayed in the Protocol Panel.

Statistical information is displayed in several ways:
Mouseover Text: Placing the mouse pointer over a term of interest will display a small table with the results of statistical analysis.
Color Code: each term is colored according to its z-score. The more an item deviates from its distributions mean, the deeper the color. Green color indicates overrepresented items, while underrepresented terms are colored in red.
Filter Statistics: For each active hierarchical filter in BiblioSphere a "Filter Statistics" table is available in the "View Panel". This view component is interlinked with the corresponding filter and allows for sorting and export of data.

Gene Ontology Filter

Filters for each category of Gene Ontology are available. Integrated statistical rating allows for the identification and selection of clusters of functionally related genes.
Each GO Filter consists of a hierarchy of terms and the corresponding annotations for your BiblioSphere. This hierarchy is originally a directed acyclic graph (DAG), but for easier navigation it has been converted to a tree for the GO Filter.
To activate the GO Filter using the selected terms, click the “Filter nodes” button. Clicking the Reset Button will deactivate the GO Filter.



Hierarchical Annotations

Each node in the tree displays the term name and the number of nodes annotated with this term for your active BiblioSphere. In the example above 294 genes are either directly annotated with the selected term "signal transducer activity" or with one of its more specific terms in the subcategories of this branch.

MeSH Filter

BiblioSphere's MeSH Filter enables you to group and filter the PubMed abstracts and genes in your BiblioSphere by MeSH Annotations.

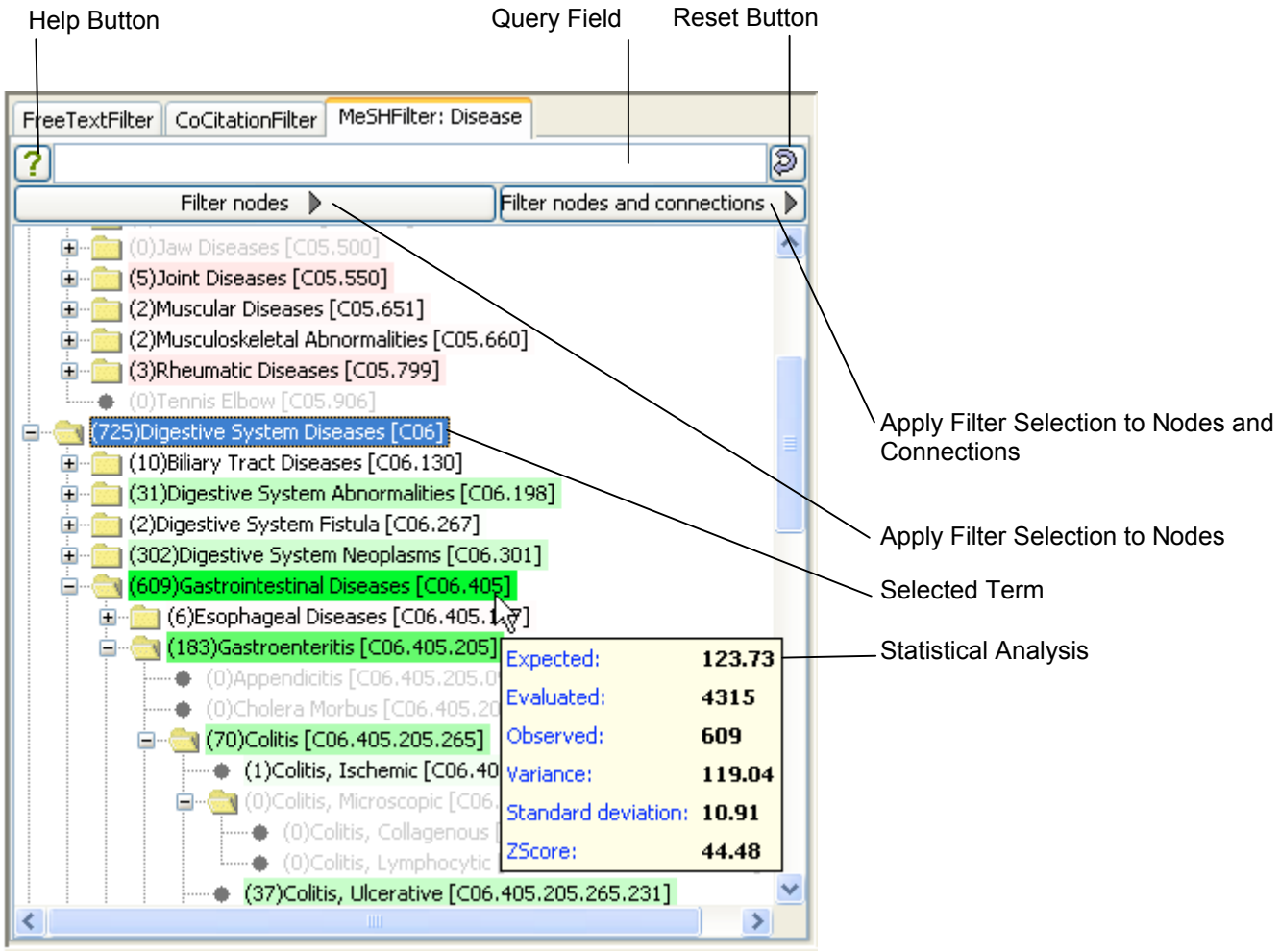
MeSH Filter Structure

Separate MeSH Filters are available for selected categories of MeSH hierarchy:

- Disease
- Chemicals and Drugs
- Anatomy
- Biological Sciences
- Analytical, Diagnostic and Therapeutic Techniques and Equipment

Each MeSH Filter consists of a hierarchy of terms and the corresponding annotations for your BiblioSphere. While GO Filters and Tissue Filters contain annotations for genes, the MeSH Filters contains this information for PubMed articles. Filtering is performed on documents. Thus, genes are filtered indirectly, as only genes co-cited in the filtered documents are displayed.

Clicking the "Filter nodes" button will apply the MeSH Filter to articles that just cite one of the genes, using the selected terms. Consequently, a connection between two genes that are co-cited in an abstract that is not annotated with the selected MeSH filter term will be displayed, if both genes appear in other accordingly annotated abstracts. In contrast, the "Filter nodes and connections" button filters for articles that contain the co-citation and are annotated with the selected term, which is more stringent. The number of abstracts can exceed the number of genes in the network due to multiple citations of the same gene. Clicking the Reset Button will deactivate the GO Filter.



Hierarchical Annotations

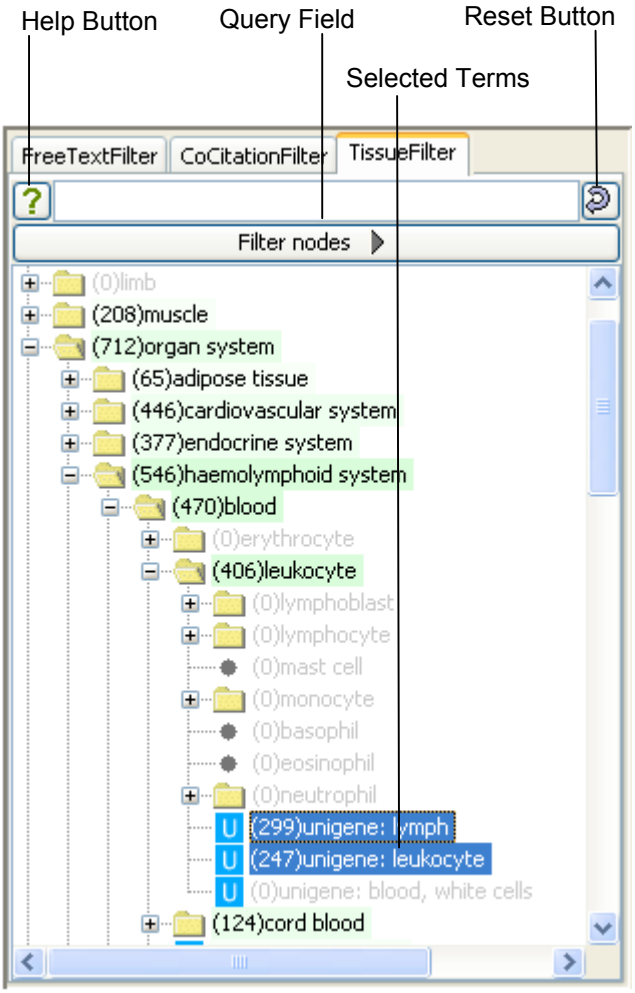
Each node in the tree displays the term name and the number of abstracts annotated with this term for your active BiblioSphere. In the example above 609 PubMed articles are either directly annotated with the selected term "Gastrointestinal Diseases" or with one of its more specific terms in the subcategories of this branch.

Tissue Filter

The Tissue Filter allows you to identify clusters of genes that share a common expression profile.

Tissue Filter Structure

Genomatix has assigned UniGene tissue names to a hierarchical tissue ontology. Thus the Genomatix hierarchical filter concept can be applied to UniGene expression data, and groups of genes with significant coexpression profiles can be identified.

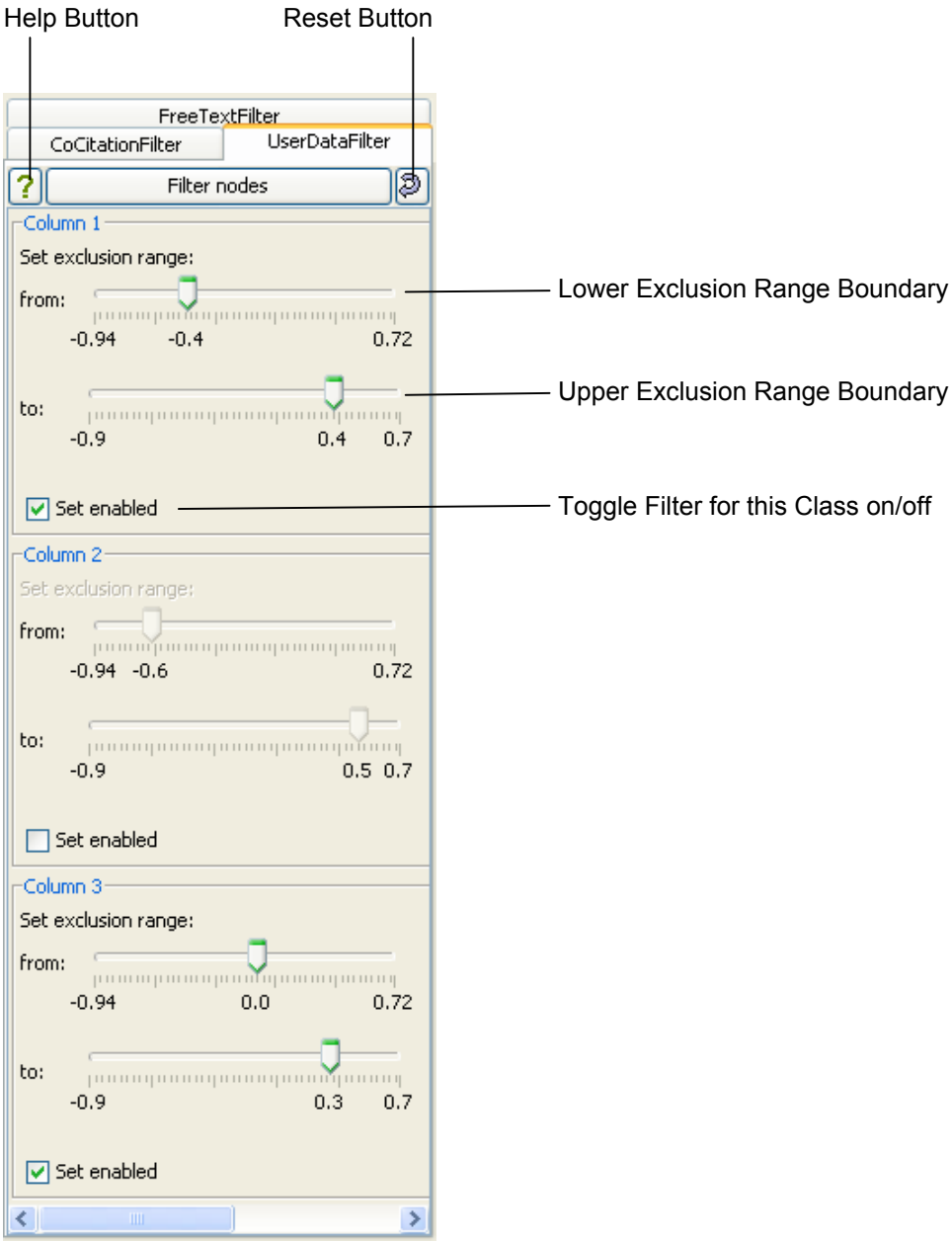


Hierarchical Annotations

Each node in the tree displays the term name and the number of genes annotated with this tissue term for your active BiblioSphere. In the example above 406 genes are annotated with UniGene tissue names assigned to the "leukocyte"-branch of the hierarchy.

User Data Filter

The User Data Filter is available if expression values were provided in the data file that was uploaded for analysis. It allows you to define an exclusion range for these values, so that only genes with an expression value below or above that range will be displayed. If expression values from a multi-class analysis were provided, you can set the range for each experimental class separately.



Sub Network Filter

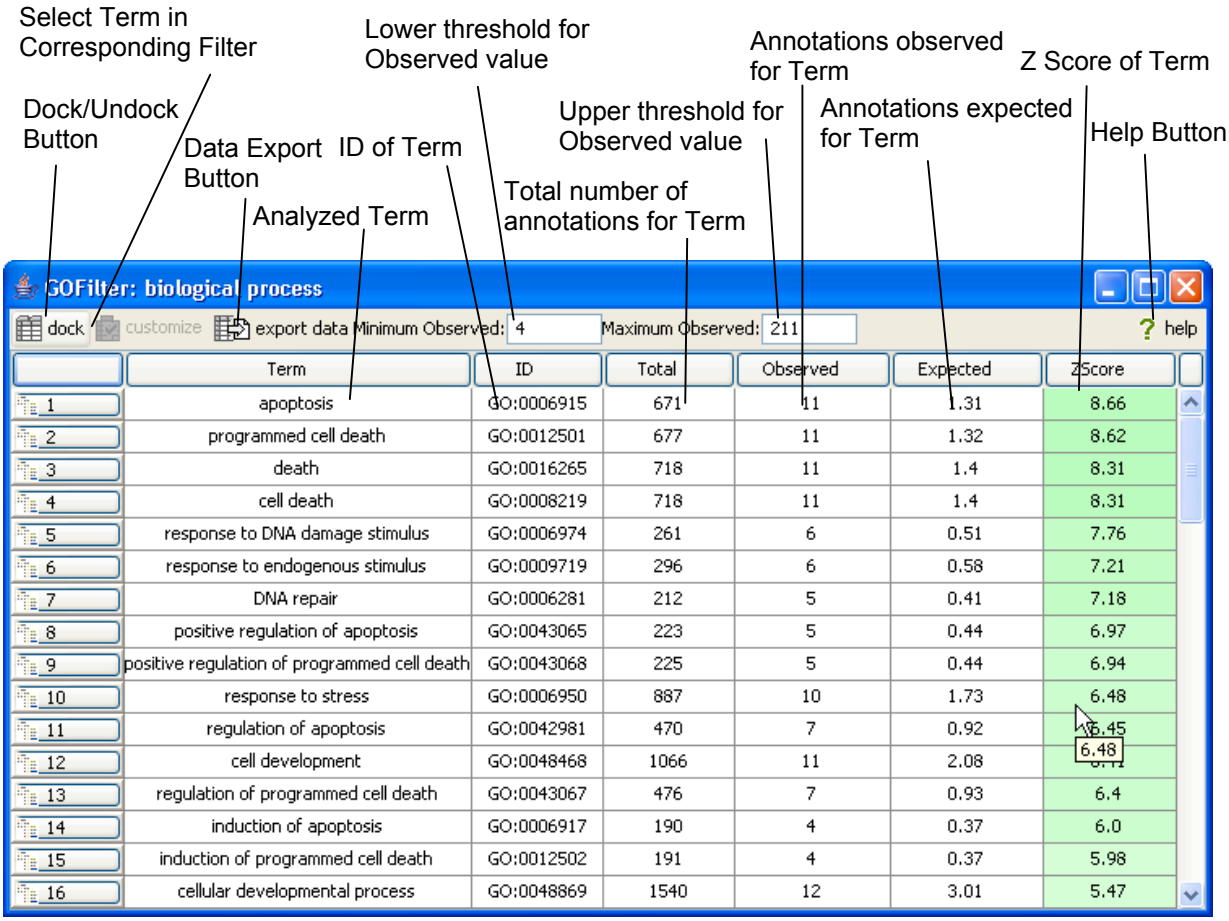
You can focus on gene subnets (clusters) by activating the Cluster View option in the BiblioSphere 3D View. This will affect other tabular and network views as well.

Statistical Analysis

Statistical Rating

Every hierarchical filter added to a BiblioSphere is statistically analyzed for over and underrepresented terms based on the number of observed and expected annotations for each term. The z-score of a term indicates whether a certain annotation or group of annotations is over- or underrepresented in the treated set. This helps you to determine whether the accumulation of annotations in a certain branch of the tree is meaningful.

To simplify the inspection of the results of this analysis, a spreadsheet view is displayed. The spreadsheet data can be exported easily to other applications (such as MS Excel) for further analysis.



	Term	ID	Total	Observed	Expected	ZScore
1	apoptosis	GO:0006915	671	11	1.31	8.66
2	programmed cell death	GO:0012501	677	11	1.32	8.62
3	death	GO:0016265	718	11	1.4	8.31
4	cell death	GO:0008219	718	11	1.4	8.31
5	response to DNA damage stimulus	GO:0006974	261	6	0.51	7.76
6	response to endogenous stimulus	GO:0009719	296	6	0.58	7.21
7	DNA repair	GO:0006281	212	5	0.41	7.18
8	positive regulation of apoptosis	GO:0043065	223	5	0.44	6.97
9	positive regulation of programmed cell death	GO:0043068	225	5	0.44	6.94
10	response to stress	GO:0006950	887	10	1.73	6.48
11	regulation of apoptosis	GO:0042981	470	7	0.92	6.45
12	cell development	GO:0048468	1066	11	2.08	6.48
13	regulation of programmed cell death	GO:0043067	476	7	0.93	6.4
14	induction of apoptosis	GO:0006917	190	4	0.37	6.0
15	induction of programmed cell death	GO:0012502	191	4	0.37	5.98
16	cellular developmental process	GO:0048869	1540	12	3.01	5.47

Total	The number of genes annotated with this term.
Observed	The number of genes in your BiblioSphere meeting the criterion.
Minimum Observed	Lower threshold for Observed value; by default =4 to avoid spurious statistical results based on small numbers.
Maximum Observed	Upper threshold for Observed value; by default = Observed value for root term in the filter; decrease to exclude the more general terms in the filter from the table.
Expected	The number of genes expected to meet the criterion based on observed values. for all co-cited genes in PubMed and your input set size.
ZScore	Over- or underrepresentation of the criterion expressed in multiples of the standard deviation.

Superimposition of Filters

You can activate any number of filters in parallel to make your selection more specific; the filter terms resulting from your selection in each single filter will be combined using the AND operator.

BiblioSphere PathwayEdition Help

Online Resources

To access the online help, click on "?" in the BSPE main menu and select "Help", or click on the help button in any of the BiblioSphere views.

Contacting Genomatix

If you encounter any problems, please contact support@genomatix.de.

Glossary

Cluster Centred BiblioSphere (CCBS)

A CCBS shows all genes connected with at least one member of your input set of genes. All second level connections in-between all genes are computed, regardless of whether an input gene is involved or not. This type of BiblioSphere is calculated when an analysis is performed.

Single Gene Centred BiblioSphere (SGBS)

A SGBS is based upon one input gene; this and all genes connected with it are shown. This type of BiblioSphere is pre-calculated and will be retrieved from the database when requested in an analysis.

Literature

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995)
MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data.
Nucleic Acids Res. 23, 4878-84
[PUBMED: 96128303]

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T (2005)
MatInspector and beyond: promoter analysis based on transcription factor binding sites
Bioinformatics 21, 2933-42
[PUBMED: 15860560]

Seifert M, Scherf M, Epple A, Werner T (2005)
Multievidence microarray mining.
Trends in Genetics 21, 553-8
[PUBMED: 16098629]

Scherf M, Epple A, Werner T (2005)
The next generation of literature analysis: Integration of genomic analysis into text mining.
Brief Bioinform. 6, 287-97